



Fundação Oswaldo Cruz. Escola Nacional de Saúde Pública

**Projeto: Modelagem de doenças transmissíveis: contexto, clima e espaço**

## **Modelagem espaço-temporal das notificações de leptospiroses em Salvador entre 1996 e 2006**

**Romy Ravines**

romy@fiocruz.br

**Marcel Quintana**

marcel@fiocruz.br

**Marilia Carvalho**

carvalho@fiocruz.br

7 de novembro de 2007

- Apresentação disponível no Wiki do LEG: <http://www.leg.ufpr.br/doku.php/pessoais:romy> -

## 1 Introdução

## 2 Análise Mensal

- Dados epidemiológicos
  - Especificação de Modelos
  - Comparação de Modelos
- Modelos de Séries de Tempo
- Modelos de Regressão
  - Análise de Componentes Principais
- Modelos de Regressão Univariados
- Modelos de Regressão Multivariados
  - Sobre el DIC

## 3 Análise Semanal

- Dados epidemiológicos
- Modelos de Séries de Tempo
- Modelos de Regressão Univariados
- Modelos de Regressão Multivariados

## 4 Análise Espacial



## Dados epidemiológicos

- Fonte: Estudo LVIGI “**Estudo de identificação de pacientes com leptospirose grave identificados no Hospital Couto Maia em Salvador-BA,**” do projeto “**Emerging Infectious Diseases and Urbanization**” (Fiocruz-Salvador & Universidade de Cornell/USA).
- Período: **20/03/1996 a 19/03/2006.**
- Número de pacientes: **1766** (Banco gerado no dia 08/02/2007).
- Critérios de inclusão dos pacientes: (1) Residente em Salvador. (2) Identificado pela vigilância ativa no HC Maia. (3) Teve alta/obito com suspeita clínica de leptospirose. (4) Identificado no período de 20/03/1996 a 19/03/2006.



Figura: Leptospirose



Figura: Salvador (BA)

## Dados Climatológicos

- Fonte: **Estação 83229 (Ondina):**  
Convencional - Operante - INMET
- Localização:
  - Latitude: 13°01'S
  - Longitude: 38°31'W
  - Altitude: 51,41 metros
- Variáveis:
  - **Precipitação (mm),**
  - **Temperatura Máxima (°C),**
  - **Temperatura Mínima (°C)**
  - **Umidade Relativa (%)**
- Período: **16/03/1996 - 25/03/2006**
- Frequência: Diária.

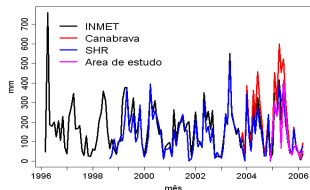
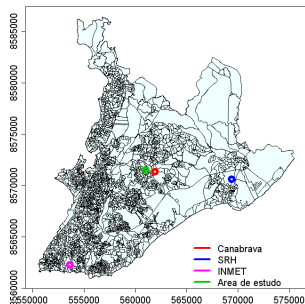


Figura: Precipitação mensal (mm)

## Notas - Tratamento dos dados

- **Período analisado:** 19/03/1996 - 25/03/2006 = 121 meses: 1996.03 - 2006.03 e 524 semanas epidemiológicas: 1996.11 – 2006.12
- **Semana Epidemiológica:** A função `as.week()` da biblioteca `epitools` do R não deu a SE correta para 1996 e 1997, o ano que lista é a do dia da metade da semana, logo não permite identificar bem os inícios e finais de ano, nem os anos com 53 semanas. Usamos uma função própria com base na data juliana e os calendários da OMS.
- **Variáveis climatológicas utilizadas:** (1) Precipitação (mm), (2) Temperatura Máxima ( $^{\circ}\text{C}$ ), (3) Temperatura Mínima ( $^{\circ}\text{C}$ ), (4) Número de Dias com Chuva, (5) Número de Dias com Precipitação > 5mm, (6) Número de Dias com Precipitação > 10mm e (7) Número de Dias com Precipitação > 15mm. Não utilizamos Umidade(%) porque a série exibe uma diminuição no final da série (pesquisando....)
- **Dados Faltantes:** Foram completados com os valores “Normais” do INMET. Testamos a opção `na.gam.replace()` da biblioteca `gam`, mas não pareceu ser uma boa opção para completar por SE.
- **Duas análises:** mensal e semanal. Considerou-se a data de internamento (sem considerar dias com sintomas).
- **Procedimento de Inferência:** Abordagem bayesiana. Métodos MCMC com o `winbugs` (R2WinBUGS do R).



## Dados epidemiológicos

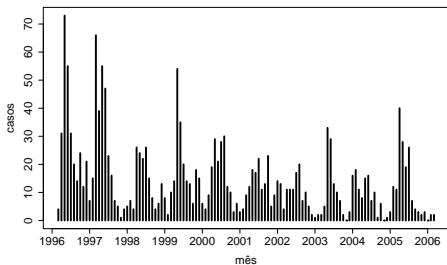
Número de casos de leptospirose grave em Salvador  
entre março de 1996 e março de 2006

Figura: Time series

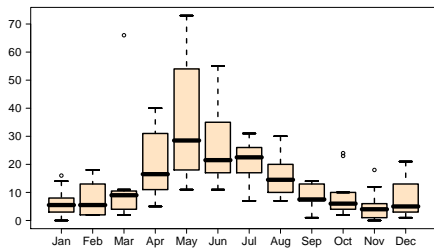


Figura: Box-Plot



# Distribuição dos dados epidemiológicos?

## Número de Casos

- Distribuição Poisson?
- Over-dispersion?
- Excesso de zeros?

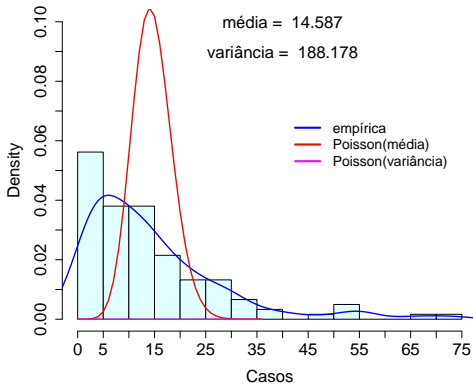


Figura: Casos mensais

## Algumas distribuições possíveis

### Poisson (*POIS*)

$$Y_t \sim \text{Poisson}(\lambda_t) \quad (1)$$

### Negative Binomial (*NB : NB01, NB02*)

$$Y_t \sim \text{Poisson}(\mu_t) \quad Y_t \sim \text{Poisson}(\lambda_t \delta_t) \quad (2a)$$

$$\mu_t \sim \text{Gamma}(\tau, \tau / \lambda_t) \quad \delta_t \sim \text{Gamma}(\epsilon, \epsilon) \quad (2b)$$

### Zero Inflated Poisson (*ZIP*)

$$Y_t \sim \pi_t \times \text{Poisson}(Y_t | \lambda_t) + (1 - \pi_t) \times \text{Poisson}(Y_t = 0 | \lambda_t) \quad (3a)$$

### Zero Inflated Negative Binomial (*ZINB*)

$$Y_t \sim \pi_t \times \text{Poisson}(Y_t | \lambda_t \delta_t) + (1 - \pi_t) \times \text{Poisson}(Y_t = 0 | \delta_t \lambda_t) \quad (4a)$$

$$\delta_t \sim \text{Gamma}(\epsilon, \epsilon) \quad (4b)$$

## Algumas especificações possíveis

- *First Order Model (FO)*

$$\log(\lambda_t) = \theta_t \quad (5a)$$

$$\theta_t = \theta_{t-1} + w_t, \quad w_t \sim N(0, W) \quad (5b)$$

- *Second Order Model (SO)*

$$\log(\lambda_t) = \theta_{1,t} \quad (6a)$$

$$\theta_{1,t} = \theta_{1,t-1} + \theta_{2,t-1} + w_{1t}, \quad w_{1t} \sim N(0, W_1) \quad (6b)$$

$$\theta_{2,t} = \theta_{2,t-1} + w_{2t}, \quad w_{2t} \sim N(0, W_2) \quad (6c)$$

- *Regression Model (X)*

$$\log(\lambda_t) = \beta_0 + \beta x_t \quad (7)$$

- *Zero inflated Model (ZI)*

$$\text{logit}(\pi_t) = \gamma_0 + \gamma z_t \quad (8)$$

# Regressão Binomial Negativa

## Parametrização I

$$y_t \sim NB\left(\tau, \frac{\tau}{\lambda_t + \tau}\right) \begin{cases} y_t \sim \text{Poisson}(\mu_t) \\ \mu_t \sim \text{Gamma}\left(\tau, \frac{\tau}{\lambda_t}\right) \end{cases} \quad (9)$$

$$\tau \sim \text{Gamma}(0.1, 0.1) \quad (10)$$

$$\log(\lambda_t) = \beta_0 + \beta_1 x_t \quad (11)$$

## Parametrização II

$$y_t \sim NB\left(\epsilon, \frac{\epsilon}{\lambda_t + \epsilon}\right) \begin{cases} y_t \sim \text{Poisson}(\lambda_t \delta_t) \\ \delta_t \sim \text{Gamma}(\epsilon, \epsilon) \end{cases} \quad (12)$$

$$\epsilon \sim \text{Gamma}(0.1, 0.1) \quad (13)$$

$$\log(\lambda_t) = \beta_0 + \beta_1 x_t \quad (14)$$

## Comparação de Modelos

- **Deviance (DEV)**

Para uma verossimilhança  $p(y|\theta)$ , o *deviance* é

$$D(\theta) = -2\log(p(y|\theta))$$

$\bar{D}(\theta)$  é uma medida de ajuste. Modelos mais complexos têm menor  $\bar{D}$ .

- **Deviance Information Criterion (DIC):**

$$DIC = \bar{D}(\theta) + pd = 2\bar{D}(\theta) + D(\bar{\theta})$$

- **Erro Quadrático Médio (EQM)**

$$EQM = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

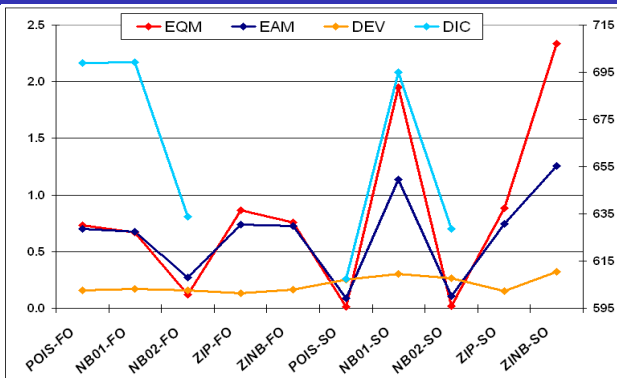
- **Erro Absoluto Médio (EAM)**

$$EAM = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

- $AIC = -2\log p(y|\hat{\psi}) + 2p_{\psi}$  (falta!)

- $BIC = -2\log p(y|\hat{\psi}) + p_{\psi} \log n$  (falta!)

## Número de Casos: MODELOS DE SÉRIES DE TEMPO



	EQM	EAM	DEV	DIC
POIS – FO	0.732	0.701	602.709	699.061
NB01 – FO	0.666	0.674	603.055	699.181
NB02 – FO	0.118	0.273	602.590	633.659
ZIP – FO	0.865	0.738	601.271	
ZINB – FO	0.755	0.727	602.896	
POIS – SO	0.014	0.089	607.017	607.347
NB01 – SO	1.951	1.133	609.622	694.886
NB02 – SO	0.019	0.107	607.595	628.632
ZIP – SO	0.880	0.743	602.189	
ZINB – SO	2.337	1.257	610.543	

## Modelo de Série de Tempo Selecionado

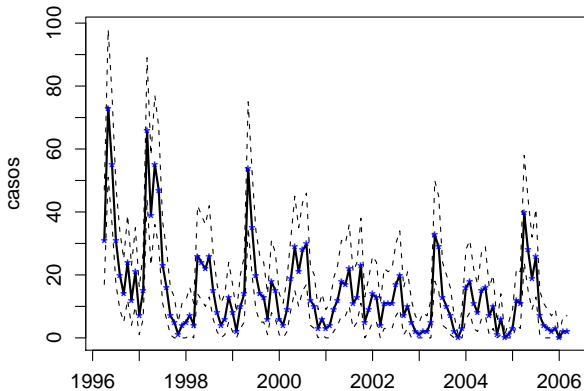


Figura: *Poisson Second Order Model*: valores replicados,  $\hat{y}_t$ .

## Modelos de Primeira Ordem

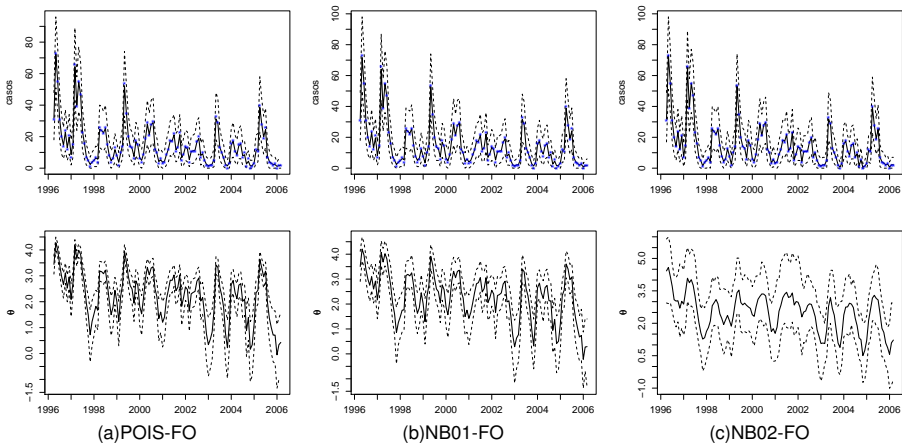


Figura: *First Order Model*: valores replicados,  $\hat{y}_t$ , e nível,  $\theta_t$ .



## Modelos de Segunda Ordem

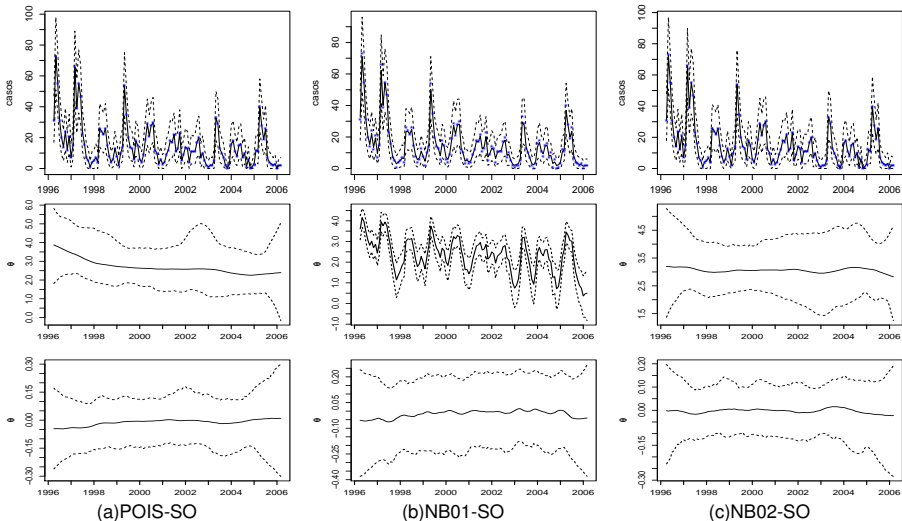
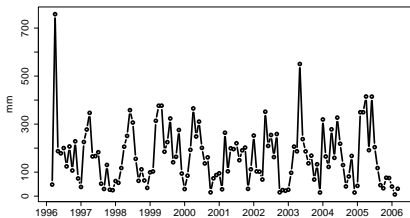


Figura: Second Order Model: valores replicados,  $\hat{y}$ , nível,  $\theta_{1t}$ , e crescimento,  $\theta_{2t}$ .

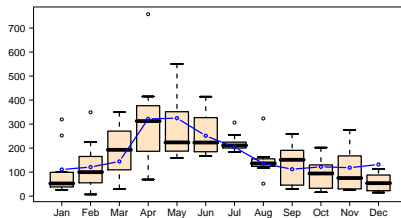
## Comentários

- Os modelos de primeira ordem não são adequados. A tendência obtida exibe uma componente sazonal (cíclica).
- Os modelos de segunda ordem são melhores. Permitem representar melhor a tendência da série. Contudo, o parâmetro de crescimento fica não significativo.
- Falta testar a inclusão de uma componente sazonal. Vamos representar a sazonalidade com alguma variável climatológica.
- Os resultados são “sensíveis” à parametrização da distribuição binomial negativa utilizada.

# Dados Climatológicos

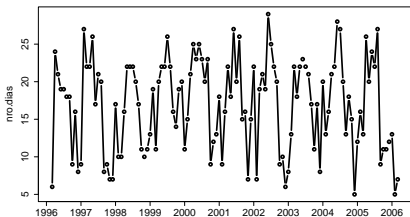


(a) Time series

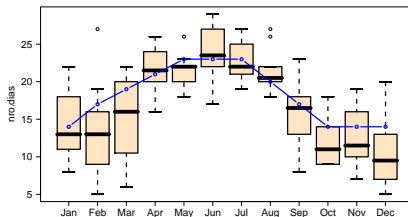


(b) Box-Plot

## Figura: Precipitação



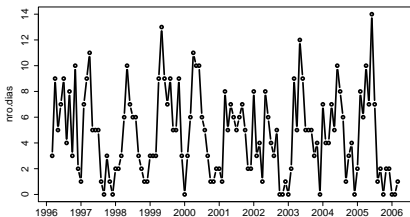
(a) Time series



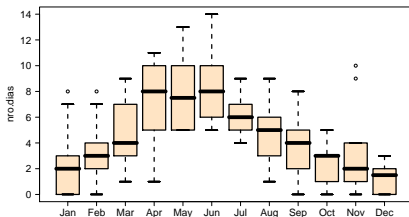
(b) Box-Plot

## Figura: Número de Dias com Chuva

## Variáveis climatológicas

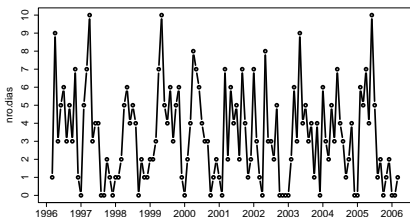


(a) Time series

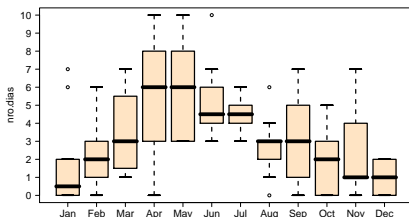


(b) Box-Plot

Figura: Dias com precipitação &gt; 10



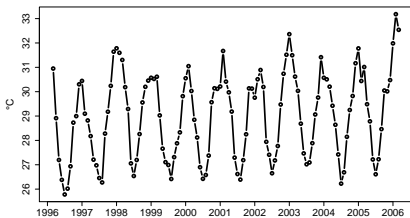
(a) Time series



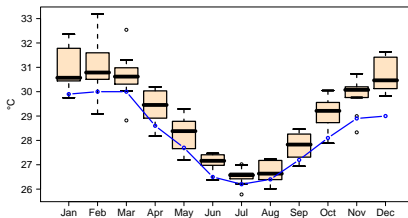
(b) Box-Plot

Figura: Dias com precipitação &gt; 15

## Variáveis climatológicas

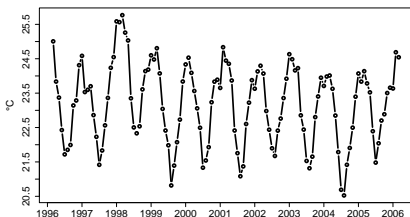


(a)Time series

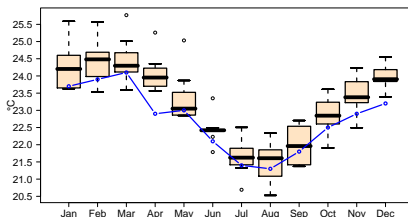


(b)Box-Plot

Figura: Temperatura Máxima



(a)Time series



(b)Box-Plot

Figura: Temperatura Mínima

# Clima: Componentes Principais

## Objetivo da Análise de Componentes Principais

Transformar um conjunto de  $p$ -variáveis  $X_1, X_2, \dots, X_p$  noutro de mesma dimensão,  $Z_1, Z_2, \dots, Z_p$ , mas com componentes ortogonais - não correlacionadas, representando diferentes dimensões nos dados.

- $Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$
- $Z_1, Z_2, \dots, Z_p$  são obtidas com as condições de que:
  - $Var(Z_1) > Var(Z_2) > \dots > Var(Z_p)$
  - $Z_i \perp Z_j \quad \forall i, j$
- Espera-se que poucas componentes principais expliquem uma grande parcela da variabilidade total.
- Se as variáveis originais forem fortemente correlacionadas, então temos, em geral, uma expressiva redução da dimensionalidade. Por exemplo, podemos ter  $p = 30$  variáveis originais expressas por apenas 2 ou 3 componentes.

# Procedimento de Componentes Principais

Os principais passos para se obter as componentes principais são:

- 1 Padronizar as variáveis  $X_1, X_2, \dots, X_p$ ;
- 2 Calcular  $C$  - a matriz de correlação;
- 3 Determinar  $\lambda_1, \lambda_2, \dots, \lambda_p$  - os autovalores de  $C$  - e obter os correspondentes autovetores; e
- 4 Descartar as componentes que expliquem pequenas proporções da variação dos dados.

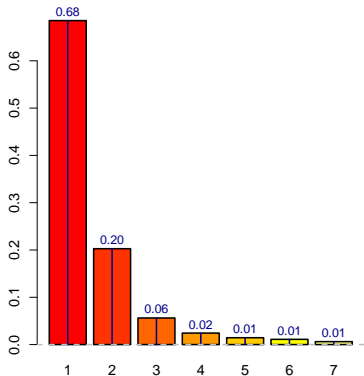
**Função princomp do R**

## Notação: variáveis

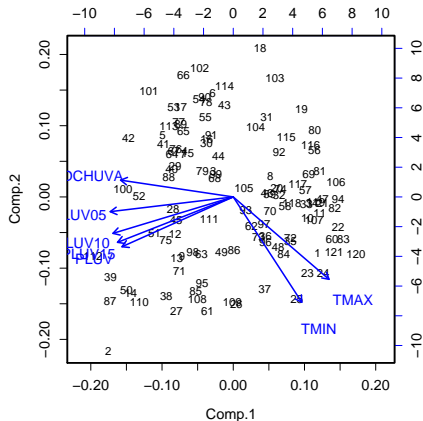
Label	Descrição
<i>PLUV</i>	Precipitação (mm)
<i>DCHUVA</i>	Nro. de dias com chuva
<i>PLUV05</i>	Nro. de dias com $PLUV \geq 5$
<i>PLUV10</i>	Nro. de dias com $PLUV \geq 10$
<i>PLUV15</i>	Nro. de dias com $PLUV \geq 15$
<i>TMAX</i>	Temperatura máxima
<i>TMIN</i>	Temperatura mínima
<i>IB1</i>	Índice de Clima 1
<i>IB2</i>	Índice de Clima 2



## Clima: Componentes Principais



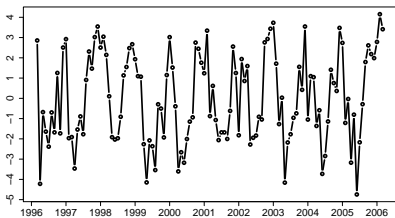
(a) Screeplot



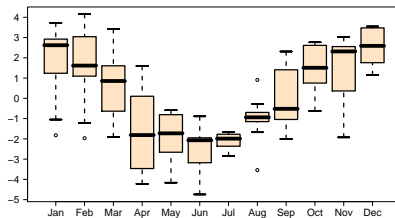
(b) Biplot

Figura: Índice de Clima

# Clima: Componentes Principais

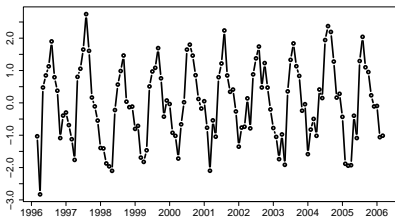


(a) Time series

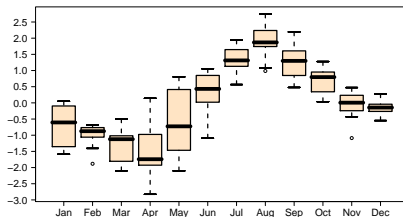


(b) Box-Plot

Figura: Índice de Clima 1

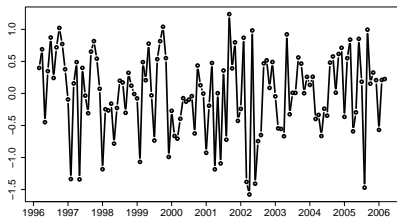


(a) Time series

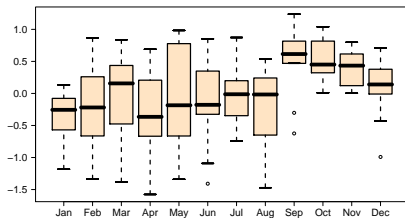


(b) Box-Plot

## Clima: Componentes Principais

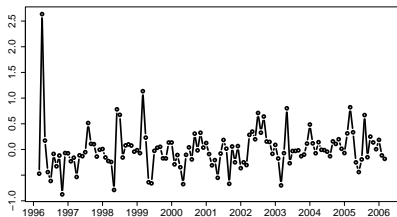


(a) Time series

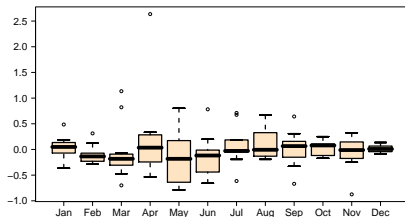


(b) Box-Plot

Figura: Índice de Clima 3



(a) Time series



(b) Box-Plot

## Clima: Componentes Principais

Correlações	<i>PLUV</i>	<i>DCHUVA</i>	<i>PLUV05</i>	<i>PLUV10</i>	<i>PLUV15</i>	<i>TMAX</i>	<i>TMIN</i>
<i>PLUV</i>	1						
<i>DCHUVA</i>	0.638	1					
<i>PLUV05</i>	0.788	0.856	1				
<i>PLUV10</i>	0.843	0.705	0.875	1			
<i>PLUV15</i>	0.883	0.607	0.822	0.939	1		
<i>TMAX</i>	-0.395	-0.666	-0.595	-0.505	-0.446	1	
<i>TMIN</i>	-0.174	-0.473	-0.370	-0.273	-0.231	0.859	1

Loadings	Comp.1	Comp.2	Comp.3	Comp.4
<i>PLUV</i>	-0.389	-0.320	0.212	<b>0.804</b>
<i>DCHUVA</i>	-0.394	0.108	<b>-0.745</b>	0.154
<i>PLUV05</i>	<b>-0.430</b>		-0.320	-0.256
<i>PLUV10</i>	<b>-0.420</b>	-0.234	0.166	<b>-0.464</b>
<i>PLUV15</i>	-0.404	-0.292	0.401	-0.207
<i>TMAX</i>	0.335	<b>-0.529</b>	-0.129	
<i>TMIN</i>	0.239	<b>-0.676</b>	-0.304	

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	2.190	1.191	0.627	0.412
Proportion of Variance	0.685	0.203	0.056	0.024
Cumulative Proportion	0.685	0.888	0.944	0.968

# MODELOS DE REGRESSÃO POISSON UNIVARIADOS

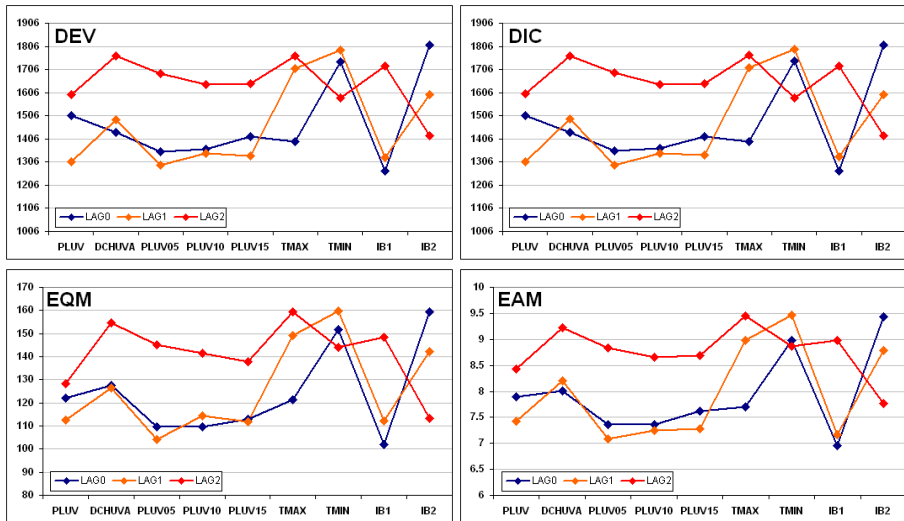


Figura: Comparação de Modelos de Regressão Poisson

# MODELOS DE REGRESSÃO BINOMIAL NEGATIVA UNIVARIADOS

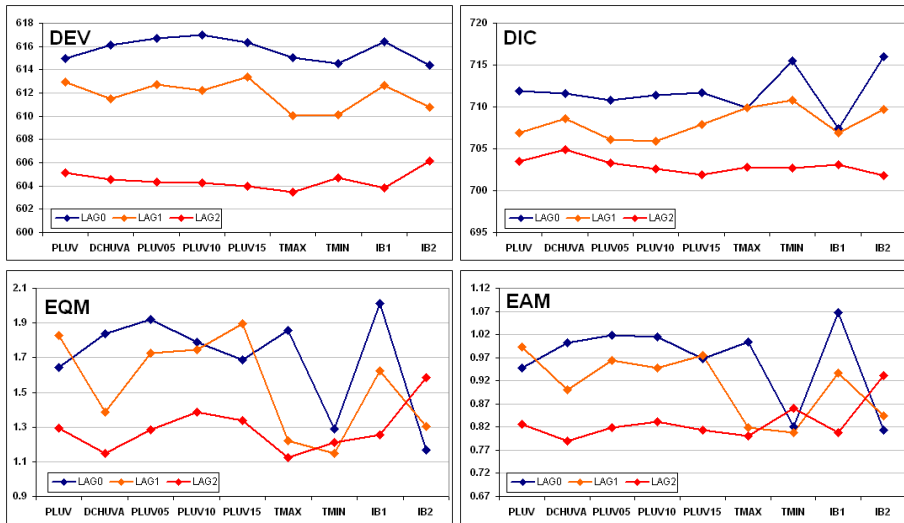


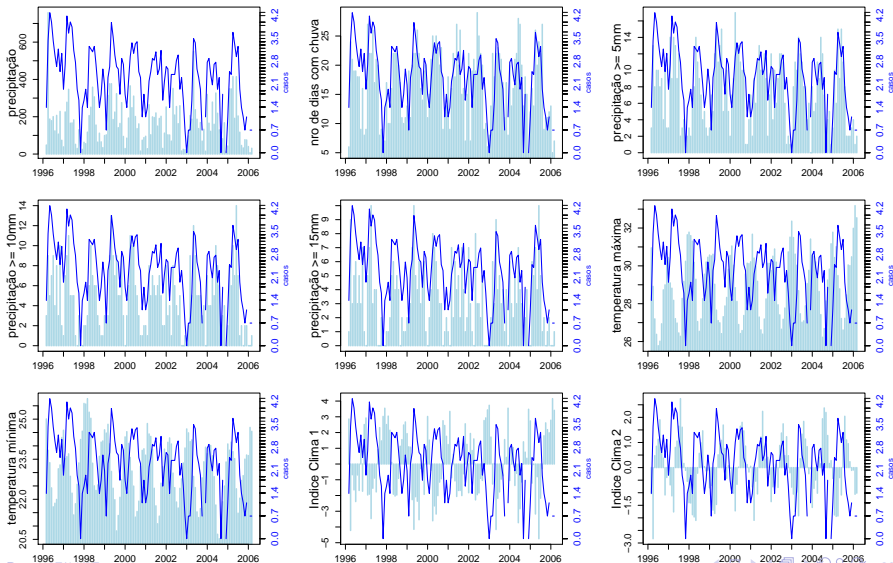
Figura: Comparação de Modelos de Regressão Binomial Negativa

# Comparação de Modelos Univariados

	POISSON						BINOMIAL NEGATIVA					
	EQM			EAM			EQM			EAM		
	LAG0	LAG1	LAG2	LAG0	LAG1	LAG2	LAG0	LAG1	LAG2	LAG0	LAG1	LAG2
<i>PLUV</i>	122.20	112.70	128.30	7.901	7.427	8.432	1.642	1.828	1.295	0.947	0.992	0.827
<i>DCHUVA</i>	127.53	126.44	154.62	8.006	8.200	9.221	1.836	1.388	1.149	1.002	0.900	0.790
<i>PLUV05</i>	109.73	103.98	145.20	7.358	7.081	8.828	1.920	1.726	1.283	1.019	0.964	0.819
<i>PLUV10</i>	109.58	114.44	141.32	7.355	7.255	8.655	1.787	1.745	1.384	1.015	0.948	0.832
<i>PLUV15</i>	112.87	111.89	137.88	7.625	7.280	8.684	1.686	1.895	1.335	0.968	0.975	0.814
<i>TMAX</i>	121.50	149.11	159.47	7.694	8.987	9.451	1.859	1.223	1.124	1.004	0.819	0.800
<i>TMIN</i>	151.57	159.58	143.91	8.986	9.468	8.874	1.291	1.147	1.212	0.820	0.808	0.861
<i>IB1</i>	101.97	112.37	148.34	6.956	7.165	8.978	2.010	1.626	1.255	1.068	0.937	0.808
<i>IB2</i>	159.33	142.35	113.15	9.436	8.778	7.769	1.169	1.304	1.583	0.814	0.845	0.931
	DEV			DIC			DEV			DIC		
<i>PLUV</i>	1505.8	1306.0	1597.7	1507.8	1308.0	1599.7	615.0	612.9	605.1	711.9	706.9	703.5
<i>DCHUVA</i>	1433.4	1489.6	1763.0	1435.4	1491.5	1765.0	616.1	611.5	604.5	711.6	708.6	704.9
<i>PLUV05</i>	1352.4	1292.5	1689.7	1354.4	1294.4	1691.6	616.7	612.8	604.3	710.8	706.1	703.3
<i>PLUV10</i>	1361.7	1341.7	1640.0	1363.6	1343.8	1642.0	617.0	612.2	604.3	711.4	705.9	702.6
<i>PLUV15</i>	1415.0	1333.0	1644.5	1417.0	1335.0	1646.5	616.3	613.4	604.0	711.7	707.9	701.9
<i>TMAX</i>	1392.6	1710.3	1764.8	1394.6	1712.5	1766.9	615.0	610.0	603.5	709.9	709.9	702.8
<i>TMIN</i>	1738.9	1790.4	1582.5	1741.0	1792.4	1584.4	614.5	610.1	604.7	715.5	710.8	702.7
<i>IB1</i>	1266.7	1325.5	1719.9	1268.7	1327.5	1721.9	616.4	612.7	603.9	707.4	706.9	703.1
<i>IB2</i>	1810.9	1596.5	1419.3	1812.9	1598.4	1421.3	614.4	610.7	606.1	716.0	709.7	701.8

- Os modelos com Binomial Negativa são “melhores” que modelos Poisson.
- Com a distribuição Poisson, os melhores modelos são de 0 ou 1 defassagens.
- Com a distribuição Binomial Negativa, os melhores modelos são os de 2 defassagens.
- Não se identifica uma “melhor” covariável. *IB1* parece ser adequada.
- Alguns problemas no processo de inferência com *TMAX* e *TMIN*.
- Foram testados modelos univariados com intercepto aleatório.

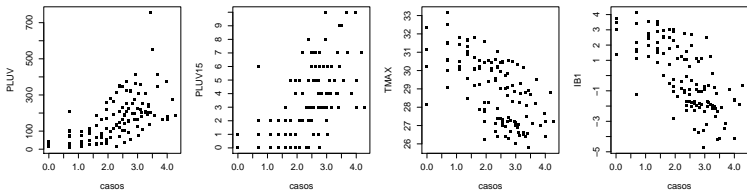
# (log)Número de casos: Relação com Clima



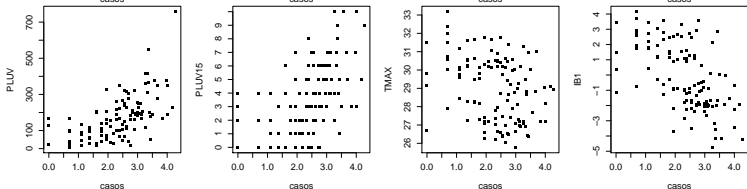


## (log)Número de casos: Relação com Clima

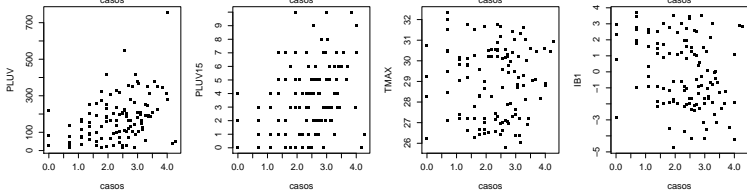
Lag 0



Lag 1



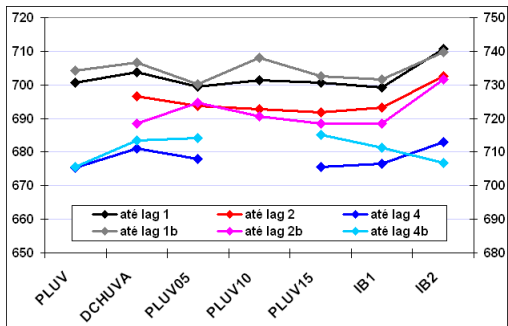
Lag 2



## Modelos Multivariados: várias defassagens

	mean	2.5%	50%	97.5%	Rhat		mean	2.5%	50%	97.5%	Rhat
PLUV 0	0.003	0.002	0.003	0.004	1.001	PLUV10 0	0.122	0.084	0.122	0.164	1.023
PLUV 1	0.004	0.003	0.004	0.005	1.001	PLUV10 1	0.103	0.060	0.102	0.147	1.010
PLUV 0	0.003	0.002	0.003	0.005	1.092	PLUV10 2	0.021	-0.019	0.021	0.061	1.000
PLUV 1	0.003	0.002	0.003	0.004	1.060	PLUV10 3	0.022	-0.018	0.022	0.065	1.002
PLUV 2	0.001	0.001	0.002	0.002	1.181	PLUV10 4	0.006	-0.030	0.006	0.044	1.001
PLUV 0	0.003	0.002	0.003	0.005	1.006	PLUV15 0	0.167	0.124	0.166	0.216	1.009
PLUV 1	0.003	0.002	0.003	0.005	1.001	PLUV15 1	0.172	0.122	0.172	0.216	1.001
PLUV 2	0.001	-0.000	0.001	0.002	1.006	PLUV15 0	0.155	0.108	0.155	0.201	1.008
PLUV 3	0.001	-0.000	0.001	0.002	1.004	PLUV15 1	0.164	0.121	0.164	0.210	1.001
PLUV 4	0.000	-0.001	0.000	0.001	1.001	PLUV15 2	0.064	0.025	0.063	0.105	1.003
DCHUVA 0	0.073	0.049	0.074	0.098	1.003	PLUV15 0	0.161	0.119	0.161	0.204	1.008
DCHUVA 1	0.050	0.026	0.050	0.072	1.085	PLUV15 1	0.145	0.100	0.145	0.192	1.002
DCHUVA 0	0.068	0.041	0.069	0.095	1.005	PLUV15 2	0.053	0.010	0.054	0.097	1.003
DCHUVA 1	0.061	0.036	0.061	0.085	1.009	PLUV15 3	0.039	-0.004	0.039	0.082	1.005
DCHUVA 2	-0.016	-0.038	-0.016	0.008	1.002	PLUV15 4	0.003	-0.035	0.003	0.042	1.001
DCHUVA 0	0.061	0.035	0.060	0.086	1.012	IB1 0	-0.221	-0.279	-0.222	-0.163	1.001
DCHUVA 1	0.058	0.032	0.057	0.084	1.007	IB1 1	-0.159	-0.214	-0.159	-0.098	1.003
DCHUVA 2	-0.004	-0.029	-0.004	0.022	1.001	IB1 0	-0.216	-0.283	-0.217	-0.151	1.001
DCHUVA 3	-0.005	-0.029	-0.005	0.020	1.002	IB1 1	-0.183	-0.251	-0.183	-0.120	1.003
DCHUVA 4	-0.013	-0.036	-0.014	0.010	1.001	IB1 2	0.044	-0.013	0.043	0.101	1.003
PLUV05 0	0.094	0.064	0.094	0.123	1.011	IB1 0	-0.203	-0.273	-0.203	-0.138	1.002
PLUV05 1	0.099	0.068	0.099	0.128	1.021	IB1 1	-0.181	-0.250	-0.182	-0.111	1.001
PLUV05 0	0.090	0.059	0.090	0.120	1.010	IB1 2	0.039	-0.032	0.039	0.105	1.001
PLUV05 1	0.106	0.075	0.106	0.135	1.001	IB1 3	0.015	-0.053	0.016	0.079	1.003
PLUV05 2	-0.008	-0.038	-0.008	0.023	1.006	IB1 4	0.015	-0.044	0.015	0.075	1.001
PLUV05 0	0.089	0.055	0.088	0.125	1.004	IB2 0	0.244	0.090	0.242	0.405	1.005
PLUV05 1	0.101	0.067	0.101	0.134	1.001	IB2 1	-0.415	-0.580	-0.417	-0.257	1.001
PLUV05 2	-0.004	-0.036	-0.005	0.030	1.005	IB2 0	0.239	0.085	0.239	0.388	1.001
PLUV05 3	-0.008	-0.039	-0.008	0.022	1.005	IB2 1	-0.208	-0.411	-0.208	-0.022	1.001
PLUV05 4	0.005	-0.028	0.004	0.036	1.001	IB2 2	-0.298	-0.445	-0.297	-0.143	1.001
PLUV10 0	0.123	0.082	0.123	0.163	1.001	IB2 0	0.022	-0.145	0.020	0.190	1.001
PLUV10 1	0.123	0.083	0.122	0.165	1.003	IB2 1	-0.182	-0.384	-0.179	-0.002	1.002
PLUV10 0	0.116	0.074	0.116	0.155	1.003	IB2 2	-0.034	-0.234	-0.037	0.168	1.001
PLUV10 1	0.117	0.076	0.117	0.159	1.004	IB2 3	-0.142	-0.345	-0.142	0.050	1.001
PLUV10 207	0.029	-0.009	0.029	0.068	1.001	IB2 4	-0.268	-0.447	-0.269	-0.089	1.001

## DIC para modelos com várias defassagens



<i>D</i>	PLUV	DCHUVA	PLUV05	PLUV10	PLUV15	IB1	IB2	PLUV	DCHUVA	PLUV05	PLUV10	PLUV15	IB1	IB2
até lag 1	613.6	611.8	612.9	613.4	614.8	613.3	612.4							
até lag 2		605.5	608.2	606.1	608.2	607.8	607.6							
até lag 4	593.0	593.0	596.7		596.9	595.9	593.0							
<i>pD</i>								<i>pV</i>						
até lag 1	87.0	92.1	86.6	88.0	85.8	85.8	98.3	120.7	124.8	117.4	124.6	117.8	118.2	127.2
até lag 2		91.0	85.6	86.7	83.6	85.4	95.0		112.9	116.6	114.6	110.4	110.8	123.9
até lag 4	82.4	88.1	81.1		78.7	80.7	89.8	112.5	120.4	117.4	120.9	118.2	115.4	113.6
<i>DIC</i>								<i>DICb</i>						
até lag 1	700.6	703.8	699.5	701.4	700.6	699.2	710.8	734.2	736.6	730.3	738.0	732.7	731.5	739.7
até lag 2		696.6	693.8	692.8	691.8	693.2	702.6		718.4	724.8	720.7	718.6	718.5	731.5
até lag 4	675.4	681.1	677.9		675.6	676.5	682.9	705.6	713.5	714.1		715.1	711.2	706.7

# Sobre el DIC

- **Deviance Information Criterion**,  $DIC = \text{'goodness of fit'} + \text{'complexity'}$

- **Ajuste** = Deviance:

$$D(\theta) = -2\log p(\text{data}|\theta)$$

- **Complexidade** = 'effective number of parameters'

$$pD = \bar{D}(\theta) - D(\hat{\theta})$$

- Logo,

$$DIC = \bar{D} + pD$$

- Com priori vaga:  $D(\theta) \approx D(\hat{\theta}) + \chi_I^2$ . Então,

$$E[D(\theta)] \approx D(\theta) + I(pD \approx I) \quad e \quad Var(D(\theta)) \approx 2I.$$

- $pV$ : alternativa ao  $pD$ :

$$pV = 1/2Var(D(\theta))$$

- "Novo" DIC:

$$DIC^* = \bar{D} + pV$$

## Modelos multivariados: várias variáveis

		mean	2.5%	50%	97.5%	Rhat
A	PLUV	0.001	-0.001	0.001	0.003	1.003
	DCHUVA	-0.003	-0.040	-0.004	0.036	1.043
	PLUV05	-0.006	-0.083	-0.004	0.064	1.022
	PLUV10	0.057	-0.047	0.059	0.155	1.003
	PLUV15	0.002	-0.139	0.001	0.147	1.017
	TMAX	-0.636	-0.793	-0.637	-0.480	1.043
	TMIN	0.630	0.426	0.633	0.839	1.049
B	IB1	-0.299	-0.367	-0.299	-0.236	1.011
	IB2	0.035	-0.074	0.035	0.142	1.002
	IB3	-0.018	-0.231	-0.021	0.197	1.002
	IB4	-0.092	-0.388	-0.096	0.225	1.002
C	PLUV	0.001	-0.001	0.001	0.003	1.002
	PLUV10	0.086	0.018	0.086	0.159	1.001
	TMAX	-0.239	-0.318	-0.238	-0.159	1.001

- As variáveis do modelo A tiveram que ser “centralizadas”.
- As variáveis do modelo B estão na sua escala original.
- Falta testar a inclusão de variáveis defassadas.

## Modelos multivariados: abordagem “hierárquica”

Tabela: Bloco I

	mean	2.5%	50%	97.5%	Rhat
DCHUVA-0	0.015	-0.012	0.019	0.048	1.009
DCHUVA-1	0.013	-0.001	0.024	0.046	1.008
PLUV10-0	0.026	0.052	0.104	0.154	1.003
PLUV10-1	0.025	0.034	0.084	0.133	1.011
PLUV-0	0.001	-0.000	0.001	0.003	1.005
PLUV-1	0.001	0.001	0.002	0.004	1.019
PLUV10-0	0.033	0.007	0.071	0.134	1.003
PLUV10-1	0.031	0.004	0.060	0.125	1.015
PLUV-1	0.001	0.001	0.002	0.004	1.013
PLUV10-0	0.020	0.083	0.121	0.161	1.010
PLUV10-1	0.032	-0.012	0.055	0.116	1.008

Tabela: Bloco II

	mean	2.5%	50%	97.5%	Rhat
TMAX-0	-0.798	-1.054	-0.791	-0.595	1.044
TMAX-1	0.107	-0.102	0.105	0.326	1.021
TMIN-0	0.640	0.320	0.631	1.008	1.018
TMIN-1	0.044	-0.288	0.053	0.333	1.022
TMAX-0	-0.784	-0.915	-0.785	-0.649	1.009
TMIN-0	0.793	0.592	0.796	0.988	1.009
TMAX-1	-0.672	-0.836	-0.670	-0.517	1.006
TMIN-1	0.914	0.694	0.908	1.173	1.005

Tabela: Bloco III

	mean	2.5%	50%	97.5%	Rhat
IB4-0	-0.097	-0.382	-0.094	0.202	1.001
IB4-1	0.133	-0.108	0.132	0.403	1.002
IB3-0	0.020	-0.242	0.021	0.294	1.001
IB3-1	-0.070	-0.314	-0.071	0.175	1.001
IB2-0	0.250	0.089	0.250	0.405	1.004
IB2-1	-0.431	-0.595	-0.434	-0.269	1.001
IB3-1	-0.186	-0.409	-0.184	0.024	1.001
IB1-0	-0.193	-0.260	-0.193	-0.127	1.005
IB1-1	-0.174	-0.263	-0.173	-0.082	1.005
IB2-0	-0.040	-0.222	-0.041	0.148	1.004
IB2-1	-0.145	-0.302	-0.144	0.002	1.002
IB1-0	-0.176	-0.240	-0.176	-0.112	1.002
IB1-1	-0.230	-0.309	-0.230	-0.157	1.002
IB2-0	-0.187	-0.302	-0.187	-0.073	1.001
IB1-0	-0.199	-0.260	-0.199	-0.141	1.001
IB1-1	-0.158	-0.216	-0.159	-0.102	1.001
IB2-1	-0.170	-0.265	-0.170	-0.078	1.002

## Modelos multivariados: abordagem “hierárquica”

Tabela: Bloco I+II+III

	mean	2.5%	50%	97.5%	Rhat
PLUV-1	0.002	0.001	0.002	0.004	1.030
PLUV10-0	-0.042	-0.182	-0.040	0.076	1.036
PLUV10-1	0.043	-0.021	0.043	0.101	1.004
TMAX-0	-0.261	-0.458	-0.258	-0.093	1.036
TMIN-0	-0.082	-0.225	-0.080	0.057	1.026
PLUV-1	0.002	0.001	0.002	0.004	1.012
PLUV10-0	0.111	0.070	0.111	0.151	1.009
PLUV10-1	0.065	0.004	0.063	0.132	1.030
IB2-1	0.087	-0.003	0.084	0.187	1.241
PLUV-1	0.002	0.001	0.002	0.004	1.006
PLUV10-1	0.040	-0.017	0.040	0.096	1.015
TMAX-0	-0.205	-0.268	-0.204	-0.144	1.001
TMIN-0	-0.048	-0.141	-0.049	0.043	1.001
PLUV-1	0.003	0.002	0.003	0.004	1.002
TMAX-0	-0.223	-0.283	-0.222	-0.168	1.002
PLUV-0	0.001	-0.001	0.001	0.003	1.002
PLUV10-0	0.086	0.018	0.086	0.159	1.001
TMAX-0	-0.239	-0.318	-0.238	-0.159	1.001

Tabela: DICs

	$\bar{D}$	$D(\theta)$	$pD$	DIC	$pV$	DICb
Bloco I						
M01	609.1	522.2	86.8	695.9	115.7	724.7
M02	610.6	525.3	85.4	696.0	122.3	733.0
M03	610.2	524.7	85.5	695.7	119.4	729.6
Bloco II						
M04	609.6	526.0	83.6	693.2	125.3	734.9
M05	609.6	525.8	83.8	693.4	116.7	726.4
M06	610.3	517.9	92.4	702.7	127.4	737.6
Bloco III						
M07	607.5	507.1	100.5	708.0	116.4	724.0
M08	606.8	506.2	100.6	707.4	122.3	729.1
M09	608.9	511.7	97.2	706.1	120.3	729.2
M10	611.2	528.2	83.1	694.3	118.7	730.0
M11	611.7	527.9	83.7	695.4	116.3	728.0
M12	610.6	528.5	82.1	692.7	116.9	727.5
Bloco I+II+III						
M13	611.1	526.2	84.9	696.1	118.4	729.6
M14	610.2	524.8	85.4	695.6	113.3	723.5
M15	611.5	527.6	83.8	695.3	115.2	726.7
M16	611.6	528.1	83.5	695.1	120.6	732.2
M17	616.2	527.2	89.0	705.2	115.8	732.0

# Modelo selecionado

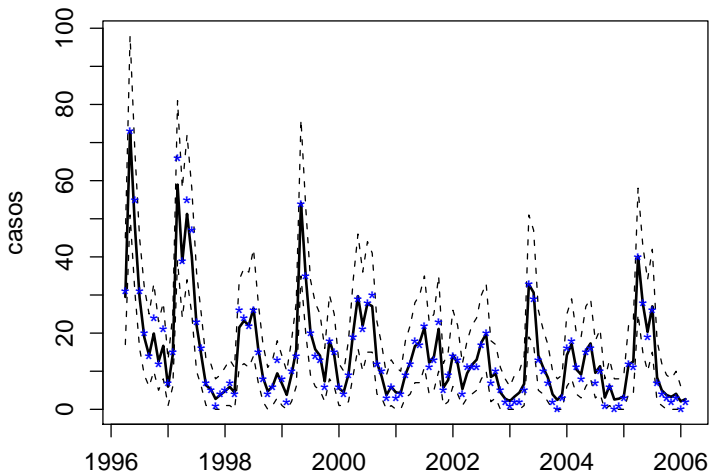
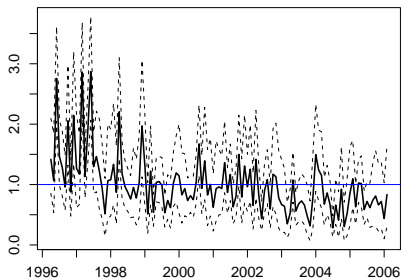


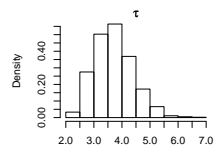
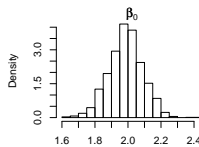
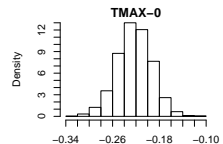
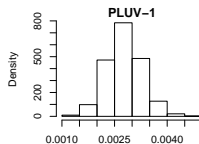
Figura: Valores replicados com modelo M16



# Modelo selecionado



(a)  $\delta_t = \mu_t / \lambda_t$

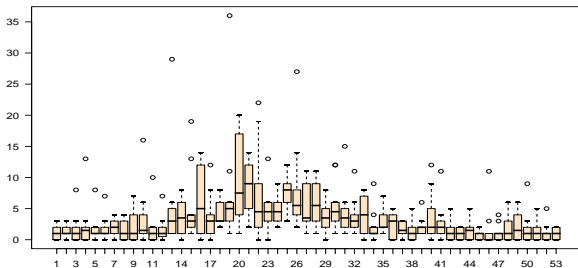
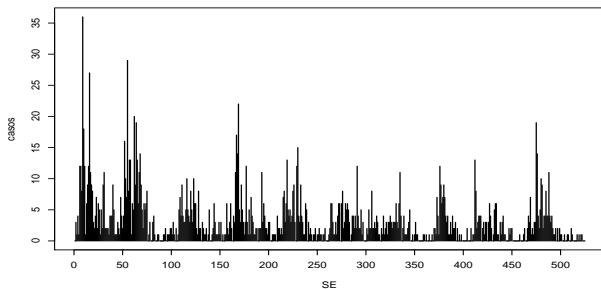


(b) Parâmetros fixos



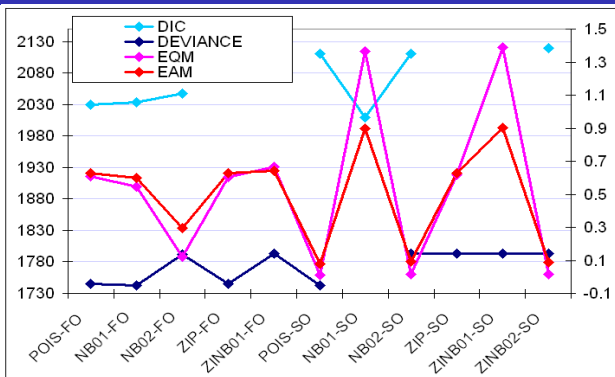
# Dados epidemiológicos

**Número de casos de leptospirose grave em Salvador entre março de 1996 e março de 2006, por semana epidemiológica**





## Número de Casos: MODELOS DE SÉRIES DE TEMPO



	EQM	EAM	DEVIANCE	DIC
POIS - FO	0.610	0.628	1745.4	2030.3
NB01 - FO	0.544	0.597	1743.3	2033.7
NB02 - FO	0.123	0.293	1792.0	2048.1
ZIP - FO	0.604	0.627	1745.7	
ZINB01 - FO	0.667	0.641	1793.2	
POIS - SO	0.012	0.081	1743.3	2111.3
NB01 - SO	1.363	0.896		2010.1
NB02 - SO	0.016	0.092	1793.2	2111.2
ZIP - SO	0.617	0.630	1793.2	
ZINB01 - SO	1.391	0.903	1793.2	
ZINB02 - SO	0.015	0.087	1793.2	2119.2









# Modelo de Série de Tempo

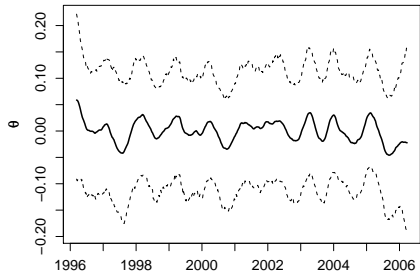
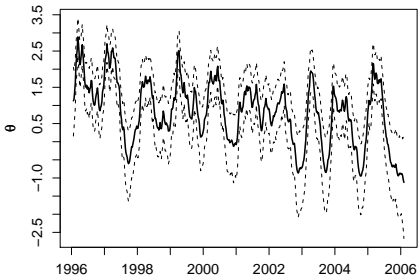
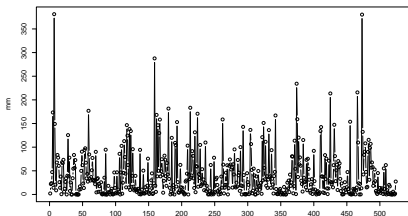


Figura: Negative Binomial Second Order Model: nível,  $\theta_1$  e crescimento  $\theta_2$ .

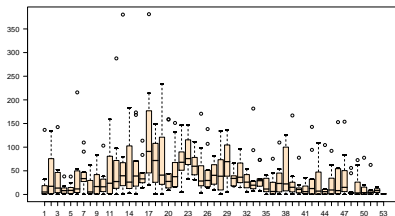
## Comentários

- Os modelos de primeira ordem têm menores valores *DIC*. A tendência obtida exibe uma componente sazonal (cíclica).
- Os modelos de segunda ordem têm menores *EAM*.
- Falta testar a inclusão de uma componente sazonal. Vamos representar a sazonalidade com alguma variável climatológica.
- Os resultados são “sensíveis” à parametrização da distribuição binomial negativa utilizada.

## Clima: Variáveis climatológicas

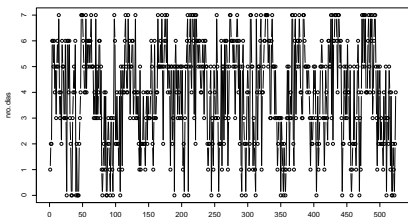


(a) Time series

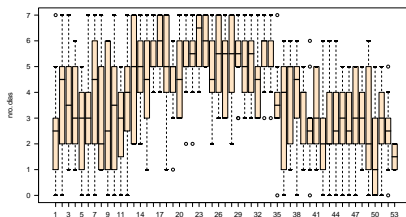


(b) Box-Plot

Figura: Precipitação



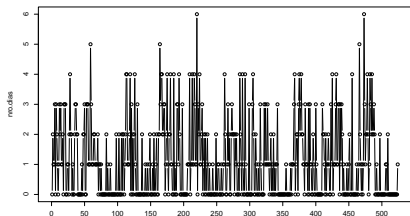
(a) Time series



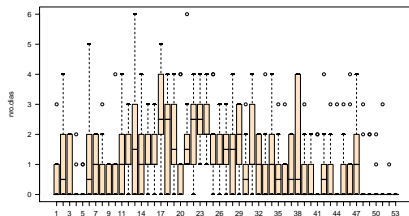
(b) Box-Plot

Figura: Número de Dias com Chuva

## Clima: Variáveis climatológicas

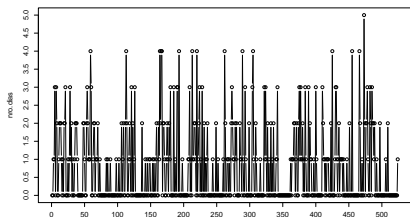


(a) Time series

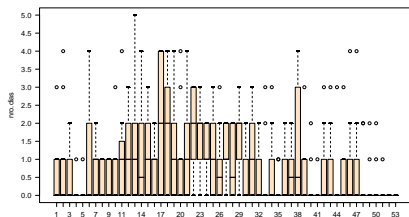


(b) Box-Plot

Figura: Dias com precipitação &gt; 10



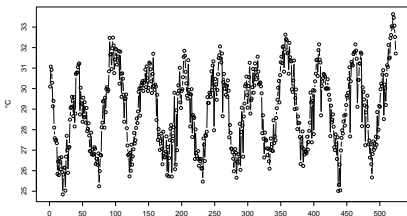
(a) Time series



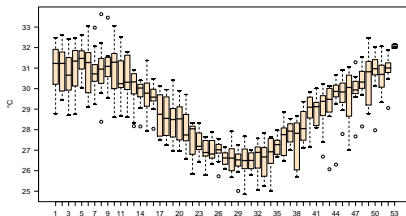
(b) Box-Plot

Figura: Dias com precipitação &gt; 15

# Clima: Variáveis climatológicas

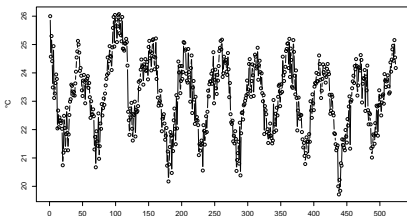


(a)Time series

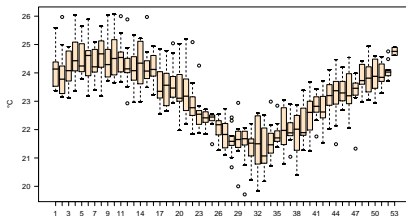


(b)Box-Plot

Figura: Temperatura Máxima



(a)Time series



(b)Box-Plot

Figura: Temperatura Mínima

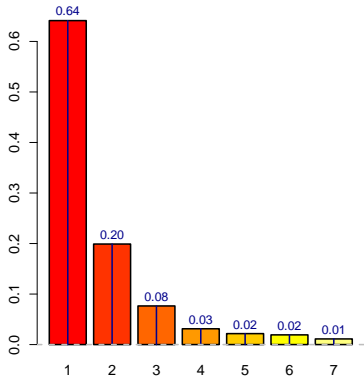
## Clima: Componentes Principais

Correlações	<i>PLUV</i>	<i>DCHUVA</i>	<i>PLUV05</i>	<i>PLUV10</i>	<i>PLUV15</i>	<i>TMAX</i>	<i>TMIN</i>
<i>PLUV</i>	1						
<i>DCHUVA</i>	0.540	1					
<i>PLUV05</i>	0.741	0.745	1				
<i>PLUV10</i>	0.811	0.591	0.842	1			
<i>PLUV15</i>	0.853	0.493	0.747	0.894	1		
<i>TMAX</i>	-0.379	-0.565	-0.508	-0.444	-0.396	1	
<i>TMIN</i>	-0.227	-0.425	-0.368	-0.295	-0.251	0.838	1

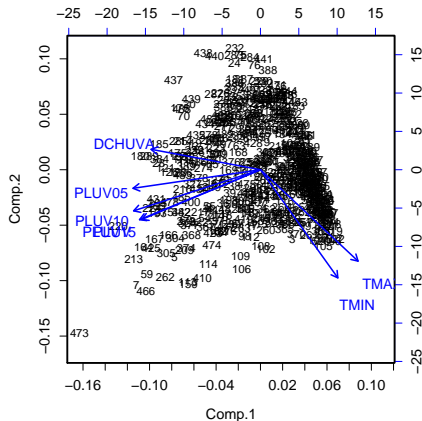
Loadings	Comp.1	Comp.2	Comp.3	Comp.4
<i>PLUV</i>	-0.399	-0.300	0.210	<b>0.699</b>
<i>DCHUVA</i>	-0.368	0.123	<b>-0.791</b>	0.274
<i>PLUV05</i>	<b>-0.428</b>	-0.112	-0.283	-0.485
<i>PLUV10</i>	<b>-0.426</b>	-0.248	0.138	-0.429
<i>PLUV15</i>	-0.406	-0.301	0.357	
<i>TMAX</i>	0.328	<b>-0.552</b>	-0.157	-0.112
<i>TMIN</i>	0.260	<b>-0.653</b>	-0.282	

Importance:	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	2.119	1.180	0.732	0.468
Proportion of Variance	0.641	0.199	0.076	0.031
Cumulative Proportion	0.641	0.840	0.917	0.948

## Clima: Componentes Principais



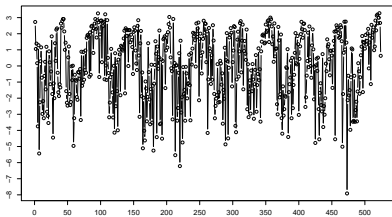
(a) Screeplot



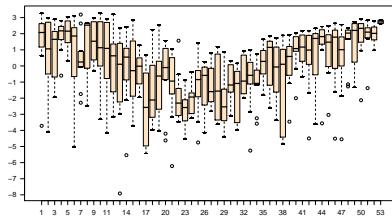
(b) Biplot

Figura: Índice de Clima

## Clima: Componentes Principais

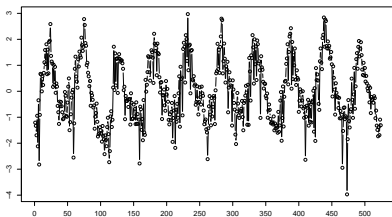


(a) Time series

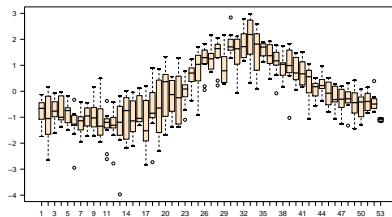


(b) Box-Plot

Figura: Índice de Clima 1



(a) Time series

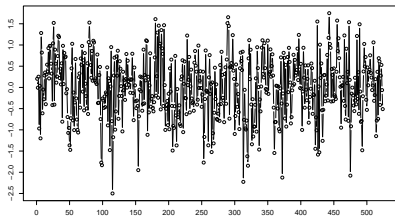


(b) Box-Plot

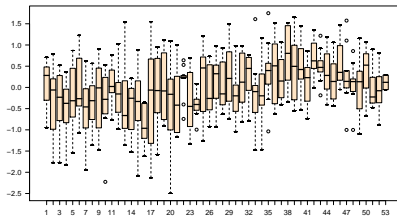
Figura: Índice de Clima 2



# Clima: Componentes Principais

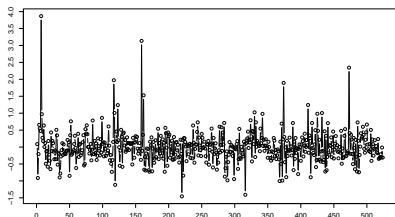


(a) Time series

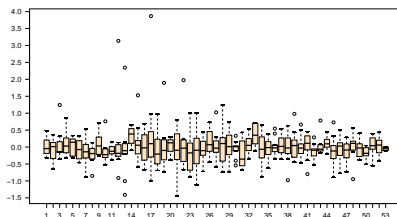


(b) Box-Plot

Figura: Índice de Clima 3



(a) Time series

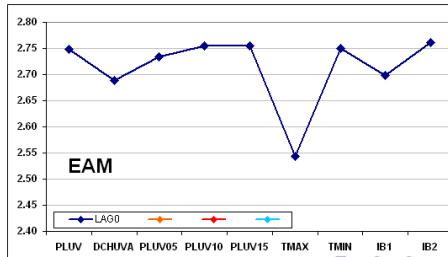
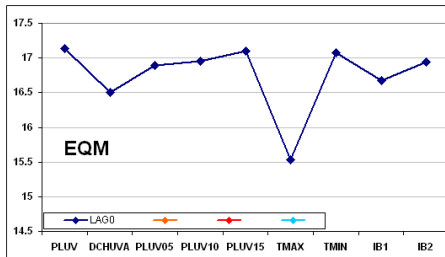
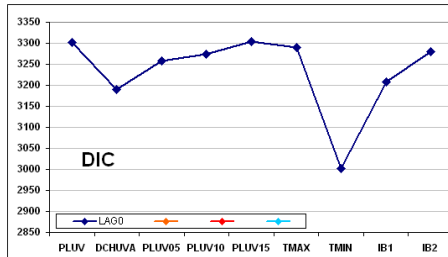
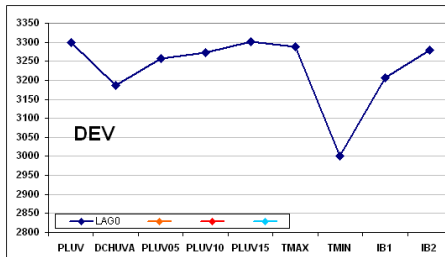


(b) Box-Plot

Figura: Índice de Clima 4

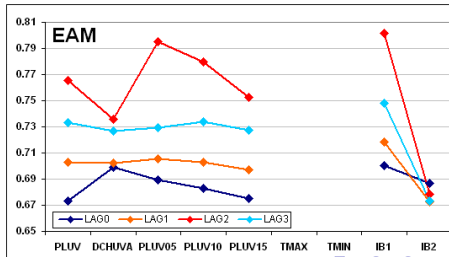
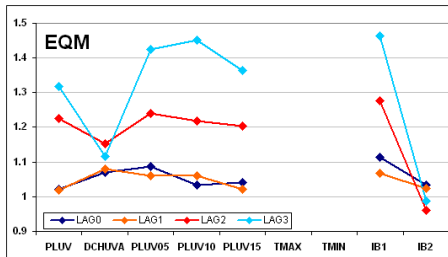
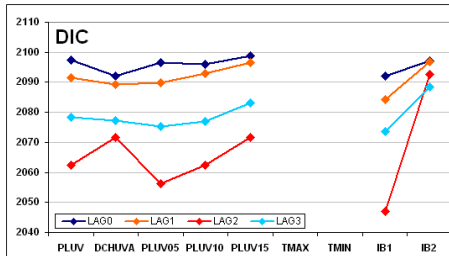
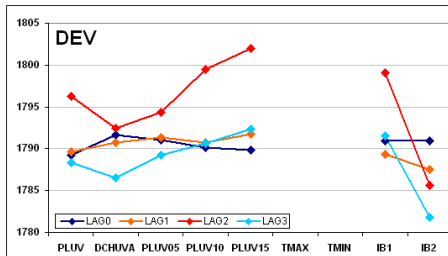
# Número de casos:

## MODELOS DE REGRESSÃO POISSON UNIVARIADOS



## Número de casos:

## MODELOS DE REGRESSÃO BINOMIAL NEGATIVA UNIVARIADOS

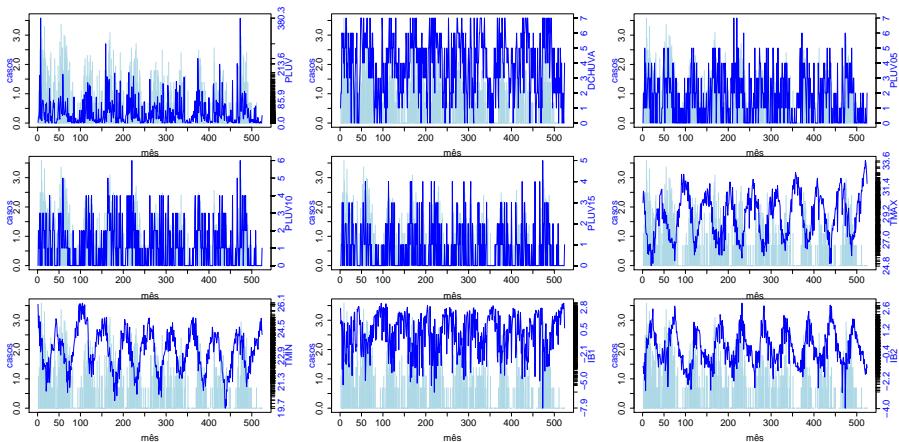


# Comparação de Modelos Univariados

	Poisson		Binomial Negativa							
	EQM	EAM	EQM				EAM			
	LAG0	LAG0	LAG0	LAG1	LAG2	LAG3	LAG0	LAG1	LAG2	LAG3
<i>PLUV</i>	17.130	2.748	1.021	1.020	1.225	1.318	0.673	0.703	0.766	0.733
<i>DCHUVA</i>	16.500	2.689	1.070	1.081	1.152	1.116	0.699	0.702	0.736	0.727
<i>PLUV05</i>	16.897	2.734	1.087	1.060	1.239	1.424	0.689	0.705	0.795	0.729
<i>PLUV10</i>	16.956	2.755	1.034	1.060	1.218	1.450	0.683	0.703	0.780	0.734
<i>PLUV15</i>	17.097	2.754	1.040	1.022	1.204	1.364	0.675	0.697	0.752	0.728
<i>TMAX</i>	15.528	2.544								
<i>TMIN</i>	17.077	2.751								
<i>IB1</i>	16.674	2.698	1.113	1.067	1.277	1.464	0.700	0.718	0.802	0.748
<i>IB2</i>	16.941	2.762	1.033	1.024	0.962	0.988	0.687	0.673	0.679	0.673
	DEV	DIC	DEV				DIC			
	LAG0	LAG0	LAG0	LAG1	LAG2	LAG3	LAG0	LAG1	LAG2	LAG3
<i>PLUV</i>	3299.2	3301.1	1789.2	1789.6	1796.3	1788.3	2097.4	2091.6	2062.4	2078.5
<i>DCHUVA</i>	3187.3	3189.3	1791.6	1790.7	1792.5	1786.6	2092.0	2089.4	2071.5	2077.3
<i>PLUV05</i>	3257.0	3259.0	1791.1	1791.3	1794.4	1789.2	2096.6	2089.9	2056.2	2075.2
<i>PLUV10</i>	3272.2	3274.3	1790.1	1790.8	1799.5	1790.6	2095.9	2093.0	2062.5	2076.9
<i>PLUV15</i>	3302.3	3304.2	1789.9	1791.8	1802.0	1792.3	2098.7	2096.7	2071.7	2083.0
<i>TMAX</i>	3287.2	3289.1								
<i>TMIN</i>	2999.9	3002.1								
<i>IB1</i>	3205.4	3207.4	1791.0	1789.3	1799.1	1791.5	2092.0	2084.2	2046.9	2073.6
<i>IB2</i>	3278.3	3280.3	1790.9	1787.5	1785.7	1781.8	2097.2	2096.8	2092.6	2088.4

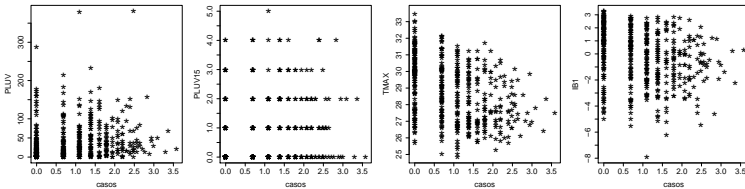
- Os modelos com Binomial Negativa são “melhores” que modelos Poisson.
- Com a distribuição Binomial Negativa, os melhores modelos são os de 2 defassagens.
- As “melhores” covariáveis são *PLUV* e *IB2*.
- Alguns problemas no processo de inferência com *TMAX* e *TMIN*.

# (log)Número de casos: Relação com Clima

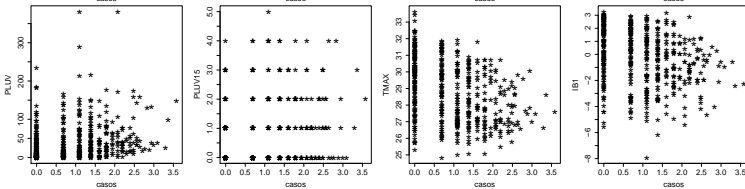


## (log)Número de casos: Relação com Clima

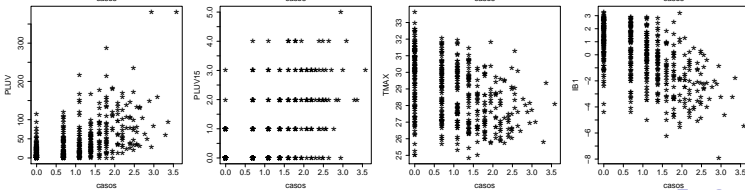
Lag 0



Lag 1



Lag 2

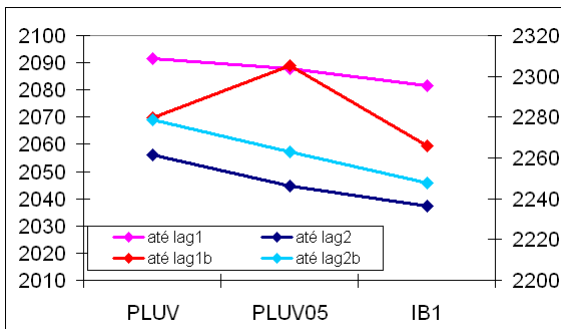


## Modelos Multivariados: várias defassagens

	mean	2.5%	50%	97.5%	Rhat
IB1-0	-0.097	-0.146	-0.097	-0.050	1.000
IB1-1	-0.156	-0.206	-0.156	-0.106	1.000
PLUV-0	0.003	0.001	0.003	0.005	1.005
PLUV-1	0.005	0.003	0.005	0.008	1.003
PLUV05-0	0.108	0.045	0.107	0.170	1.005
PLUV05-1	0.196	0.133	0.195	0.258	1.001
IB1-0	-0.064	-0.104	-0.064	-0.023	1.001
IB1-1	-0.062	-0.106	-0.062	-0.018	1.001
IB1-2	-0.268	-0.308	-0.267	-0.229	1.001
PLUV-0	0.002	0.000	0.002	0.004	1.004
PLUV-1	0.003	0.001	0.003	0.005	1.001
PLUV-2	0.011	0.009	0.011	0.013	1.013
PLUV05-0	0.071	0.017	0.071	0.124	1.003
PLUV05-1	0.119	0.063	0.118	0.174	1.002
PLUV05-2	0.342	0.287	0.342	0.394	1.012

- Todos os parâmetros são significativos.
- Testar mais defassagens.

## DIC para modelos com várias defassagens



<i>D</i>	<i>PLUV</i>	<i>PLUV05</i>	<i>IB1</i>	<i>PLUV</i>	<i>PLUV05</i>	<i>IB1</i>
até lag1	1790.2	1790.7	1789.9			
até lag2	1794.2	1789.9	1794.5			
<i>pD</i>				<i>pV</i>		
até lag1	301.3	297.0	291.7	489.5	514.6	476.2
até lag2	262.0	254.7	242.9	484.5	473.1	453.3
<i>DIC</i>				<i>DICb</i>		
até lag1	2091.5	2087.7	2081.7	2279.7	2305.3	2266.1
até lag2	2056.1	2044.6	2037.4	2278.7	2263.0	2247.8



## Modelo selecionado\*

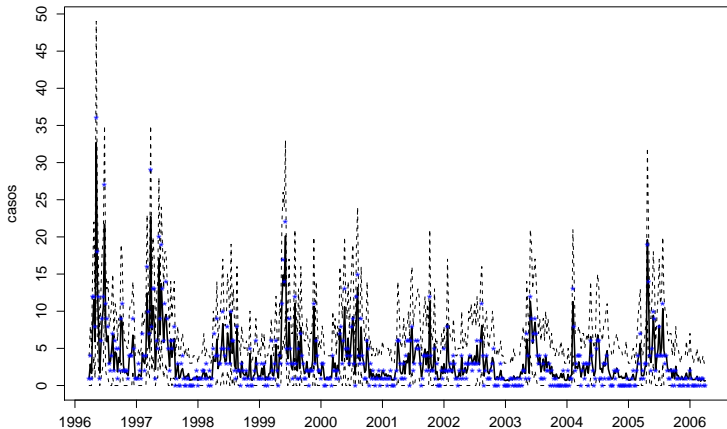


Figura: Valores replicados com modelo Binomial Negativos, com  $IB1(0)$ ,  $IB1(-1)$  e  $(IB2(-2))$ .

## Modelo selecionado\*

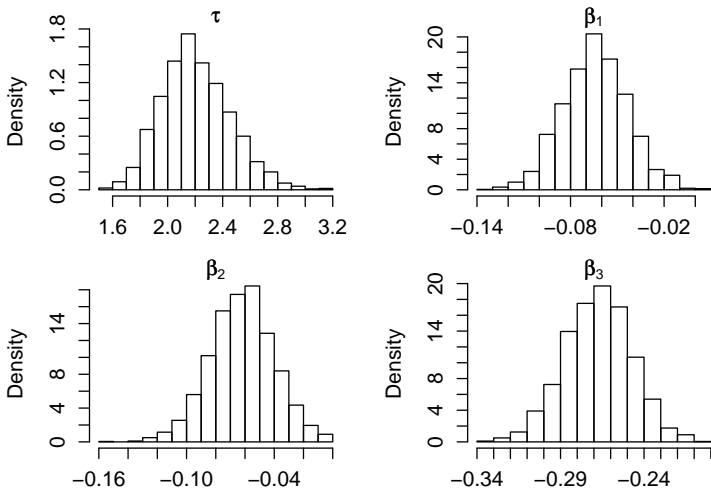


Figura: Efeitos Fixos no modelo Binomial Negativos, com  $IB1(0)$ ,  $IB1(-1)$  e  $(IB2(-2))$ .

# Comparação de Deviances

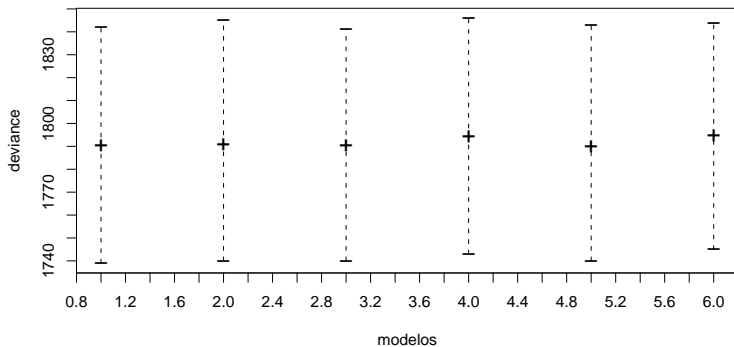


Figura: Amostras dos deviance dos modelos com várias defassagens

## **Análise Espacial**

### **Dados Sócio-econômicos por Setor Censitário**

## Regionalização de áreas: o **Skater**

- **Objetivo:** Agrupar áreas menores (municípios, bairros, setores) em áreas maiores, que além de serem contíguas, tenham também semelhança segundo os atributos associados.

### O método **SKATER**

Análise de conglomerados que leva em conta a localização espacial dos objetos. “Trata-se de uma situação em que temos um mapa particionado em áreas, cada uma delas com uma posição geográfica determinada, e desejamos que áreas com características semelhantes pertençam ao mesmo conglomerado espacial. Estas características são dadas pelas variáveis medidas”

- O método **Skater** está implementado en el **Terraview**.

## Regionalização de áreas: o Skater

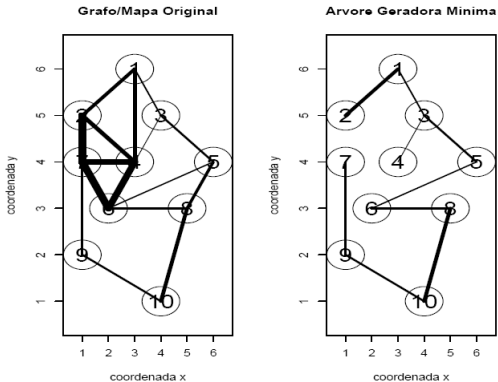


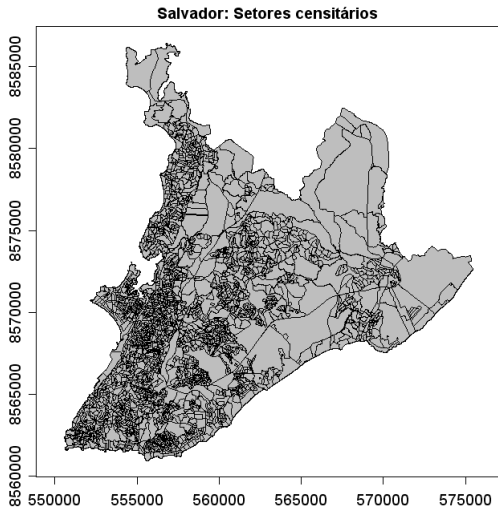
Figura 3: Grafo de mapa com 10 áreas com arestas proporcionais ao seu custo e sua árvore geradora mínima.

## Variáveis disponíveis

## Características dos setores censitários: Censo 2000.

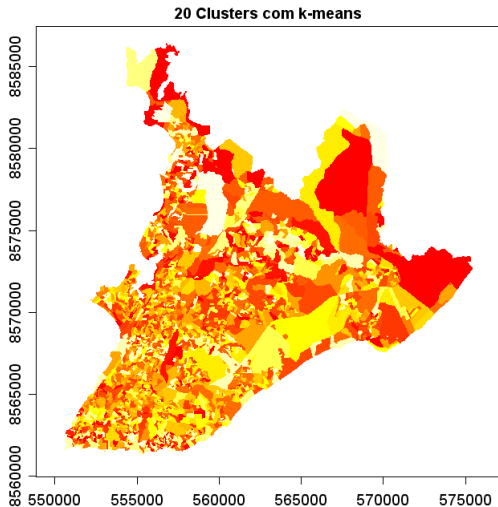
Domicílio	% dpp com abastecimento de água da rede geral
	% dpp com abastecimento de água de poço ou nascente (na propriedade)
	% dpp com abastecimento de água de outra forma (nem rede geral nem poço ou nascente)
	% dpp com banheiro ou sanitário
	% dpp sem banheiro ou sanitário
	% dpp com banheiro ou sanitário - esgotamento sanitário - rede geral de esgoto ou pluvial
	% dpp com banheiro ou sanitário - esgotamento sanitário - fossa séptica ou rudimentar
	% dpp com banheiro ou sanitário - esgotamento sanitário - vala, rio, lago, mar ou outro
	% dpp com destino do lixo coletado
	Chefe
% dpp com homens responsáveis (chefes)	
% dpp com pessoas responsáveis alfabetizadas	
% dpp com homens responsáveis alfabetizados	
% dpp com mulheres responsáveis alfabetizadas	
% dpp com pessoas responsáveis não alfabetizadas	
% dpp com homens responsáveis não alfabetizados	
% dpp com mulheres responsáveis não alfabetizadas	
% dpp com pessoas responsáveis com até 1 ano de estudo	
% dpp com homens responsáveis com até 1 ano de estudo	
% dpp com mulheres responsáveis com até 1 ano de estudo	
% dpp com pessoas responsáveis com até 1 salário mínimo	
Anos de estudo da pessoas responsáveis pelos domicílios particulares permanentes	
Rendimento nominal mensal das pessoas responsáveis pelos dpp	
Residentes	Pessoas residentes nos dpp
	Homens residentes nos dpp
	Mulheres residentes nos dpp
	% de mulheres residentes nos dpp
	% de pessoas residentes com até 5 anos de idade
	% de pessoas residentes com até 4 anos de idade
	% de pessoas residentes alfabetizadas
	% de pessoas residentes não alfabetizadas
dpp=domicílios particulares permanentes	

## Regionalização de áreas

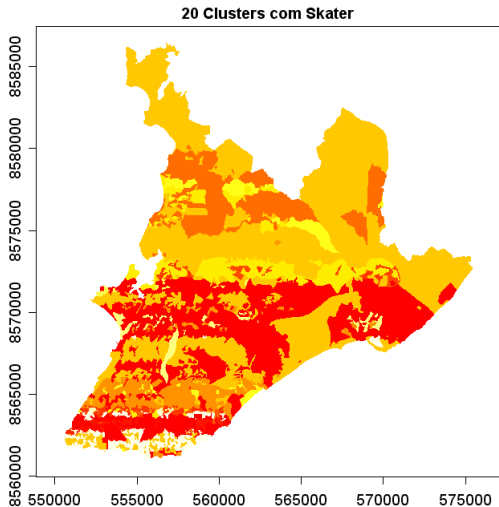




## Regionalização de áreas



## Regionalização de áreas



## Regionalização de áreas: o Skater

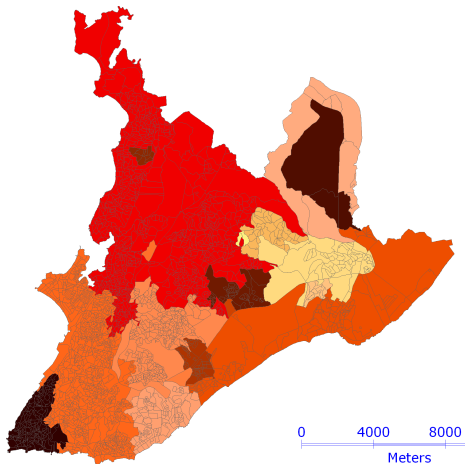


Figura: Skater no Terraview: todas as variáveis

## Regionalização de áreas: o Skater

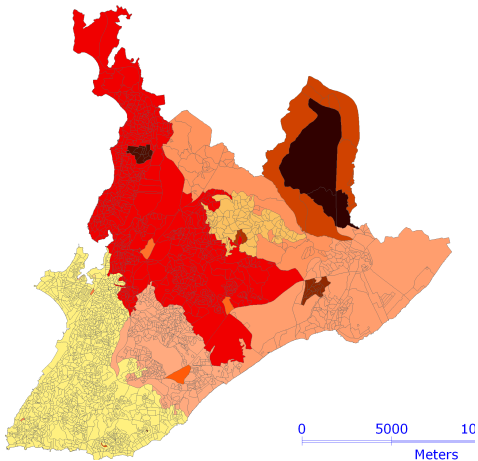


Figura: Skater no Terraview: somente variáveis sobre domicílio

## Regionalização de áreas

- Número de setores em cada cluster -

Cluster	1	2	3	4	5	6	7	8	9	10
Nro Setores	747	43	2	61	13	11	120	1	275	1
Cluster	11	12	13	14	15	16	17	18	19	20
Nro Setores	1	833	1	143	1	45	10	45	3	160
Cluster	1	2	3	4	5	6	7	8	9	10
Nro Setores	167	67	4	127	247	103	254	36	27	182
Cluster	11	12	13	14	15	16	17	18	19	20
Nro Setores	165	149	106	104	199	7	116	145	233	58