

Investigando fatores associados a contagens de ovos de *Aedes aegypti* em ovitrampas em Recife/PE

Wagner Hugo Bonat
Henrique Silva Dallazuanna
Orientador: PhD Paulo Justiniano Ribeiro Júnior

Departamento de Estatística
Setor de Ciências Exatas
Universidade Federal do Paraná

Introdução

- Aspectos sócio-culturais afetam o cenário epidemiológico brasileiro.
- Epidemias de dengue, leptospirose, tuberculose entre outras.
- Modelar e identificar fatores de risco / proteção nas situações endêmicas e epidêmicas.
- Projeto SAUDAVEL
- Desenvolver instrumentos para a prática da vigilância epidemiológica e entomológica.

Epidemiologia vs Entomologia

- **Epidemiologia** - Estuda quantitativamente a distribuição dos fenômenos de saúde/doenças e seus fatores condicionantes e determinantes nas populações humanas.

- **Entomologia** - Estuda os insetos sob todos os aspectos e relações com o homem, os animais e o ambiente.

Cenário

- A dengue é uma das mais importantes doenças da atualidade.
- Causador é um vírus, mais seus transmissores são mosquitos do gênero *Aedes*.
- Nos países de clima tropical, as condições ambientais favorecem o desenvolvimento e a proliferação.

Aedes aegypti

- É o mais comum do gênero *Aedes*.
- Domesticado e urbano se alimenta durante o dia.



Figura: Mosquito *Aedes aegypti*.

Transmissão

- A partir da picada das fêmeas adultas do mosquito.
- Somente elas são hematófogas.
- De 8 a 10 dias após ter sugado o sangue de uma pessoa contaminada, a fêmea infectada está apta a transmitir o vírus.

Dengue no Brasil

- Em 1980 iniciou-se um processo de intensa circulação do vírus da dengue.
- Epidemias explosivas em todas as regiões brasileiras.
- Não é transmissível de pessoa para pessoa.
- Portanto a única forma efetiva de contrair dengue é através da picada do mosquito.

Vigilância

- A partir da década de 80 surgiram diversas metodologias para a vigilância do vetor.
- Mapeamento de áreas de risco.
- Relação com a vigilância epidemiológica.
- 1 Presença do vetor
- 2 Índices de infestação
- 3 Características biológicas
- 4 Eficácia de inseticidas

Vigilância entomológica

- Coleta de formas imaturas (ovo, larva e pupa).
- Proposta de Connor e Monroe, 1923.
- Densidade de *Aedes* em áreas urbanas.
- Vistoriar depósitos de água.
- Cálculo de índices de infestação.

Vigilância entomológica

- A pesquisa larvária não é um bom indicador do risco de transmissão.
- Outra opção é a coleta de mosquitos adultos. (custosa e demorada)
- Ovitrapas são destinadas a coleta de ovos.
- Em 1965 iniciou-se o uso de ovitrapas (Fay, 1965).
- Posteriormente ficou demonstrada a superioridade em relação a pesquisa larvária (Fay, 1966).
- Fornecem dados sobre a distribuição espacial e temporal.

Objetivos

- Determinar fatores de risco e proteção associados a ocorrência de ovos do mosquito *Aedes aegypti*.
- Entender a dinâmica de proliferação do mosquito.
- Serão avaliados os seguintes aspectos:
 - 1 Condições ligadas as armadilhas.
 - 2 Condições climáticas.
 - 3 Relações espaciais.
 - 4 Relações temporais.

Área de estudo

- Desenvolvido em Recife/PE.
- Instaladas 564 armadilhas (ovitrampas).
- Início do monitoramento em maio de 2004.
- A cada 7 dias é feita a contagem em cerca de um quarto das armadilhas, ciclos de 28 dias são necessários para que todas as armadilhas sejam monitoradas.
- As armadilhas cobrem a superfície do bairro.
- Até maio de 2007 foram realizadas 19.068 observações, contados 14.829.557 ovos.

Aspectos Computacionais

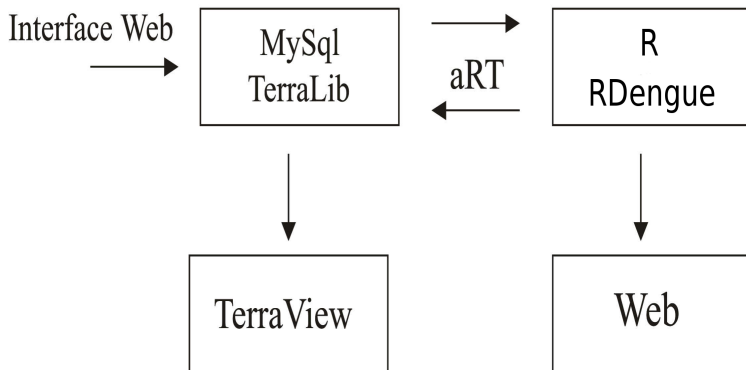


Figura: Formato geral de análise.

Construção das Covariáveis

- O experimento de Recife/PE permitiu a construção de várias covariáveis.
- Os agrupamentos foram realizados buscando dar uma interpretação prática.
- As covariáveis foram classificadas em dois grupos:
 - 1 Covariáveis locais - condições ligadas a armadilha.
 - 2 Covariáveis ambientais - condições ligadas a fatores abióticos.

Covariáveis da armadilha

- | | | |
|-----|--------------------------|-----------------|
| 1. | Tipo de Imóvel | (0) residencial |
| 2. | Presença de Quintal | (0) apresenta |
| 3. | Água ligada a rede geral | (0) ligada |
| 4. | Freq. abastecimento | (0) diário |
| 5. | Água canalizada | (0) apresenta |
| 6. | Fatores de risco | (0) apresenta |
| 7. | Grandes sem tampa | (0) apresenta |
| 8. | Grandes com tampa | (0) apresenta |
| 9. | Pequenos sem tampa | (0) apresenta |
| 10. | Pequenos com tampa | (0) apresenta |

Covariáveis climáticas

- Precipitação média semanal.
- Umidade média semanal.
- Temperatura máxima semanal.
- Temperatura mínima semanal.
- Defasagens de 12 semanas da observação.

Metodologia

- Análise fatorial.
- Modelos para dados de contagem.
- Modelos aditivos generalizados (GAM).

Análise fatorial

- Descrever relações de covariância entre diversas variáveis.
- Resume a informação através de quantidades aleatórias não observáveis denominadas de **fatores**.
- Os fatores não são restritos a ortogonalidade.
- Exige a normalidade e a determinação a priori do número de fatores.
- Forma alternativa para acomodar no modelo um conjunto de covariáveis que se correlacionam fortemente.

Modelos para dados de contagem

- Modelo log-linear de Poisson é o mais utilizado quando se trata de dados de contagem.
- Feômeno da superdispersão.
- Suposição de distribuição de Poisson é inadequada.
- Quase-Verossimilhança. (parâmetro extra de dispersão)
- Distribuição Binomial Negativa.

Distribuição Binomial Negativa

- Suponha (x_1, \dots, x_p) variáveis explicativas.
- $Y|z$ tem média z e variância z .
- Porém Z não observável varia nas unidades amostrais com x fixo, então $E(Z) = \mu$.

$$\begin{aligned}E(Y) &= E[E(Y|Z)] = E[Z] = \mu \quad \text{e} \\ \text{Var}(Y) &= E[\text{Var}(Y|Z)] + \text{Var}[E(Y|Z)] \\ &= \mu + \text{Var}(Z)\end{aligned}$$

- Suponha que $Y|z \sim \text{Poisson}$ de média μ , denotada por $f(y|z)$.
- Adicionalmente $Z \sim \text{Gamma}(\mu, k)$ onde $k = \phi\mu$, e f.d $g(z, \mu, k)$.

Distribuição Binomial Negativa

- Implica que a distribuição marginal de Y é:

$$\begin{aligned}P(Y = y) &= \int_0^{\infty} f(y|z)g(z; \mu, k)dz \\&= \frac{\Gamma(y + k)\phi^k}{\Gamma(y + 1)\Gamma(k)(1 + \phi)^{y+k}} \\&= \frac{\Gamma(y + k)}{\Gamma(y + 1)\Gamma(k)} \left(\frac{\phi}{1 + \phi}\right)^k \left(\frac{1}{1 + \phi}\right)^y \\&= \frac{\Gamma(y + k)}{\Gamma(y + 1)\Gamma(k)} (1 - \pi)^k \pi^y\end{aligned}$$

em que $\pi = \frac{1}{(1+\phi)}$, note que $Var(Z) = \frac{\mu^2}{k}$ de modo que

$Var(Y) = \mu + \frac{\mu^2}{k} = \frac{\mu(1+\phi)}{\phi}$. Portanto, $[Y]$ tem distribuição binomial negativa.

Log Verossimilhança

- $\theta = (\beta, \phi)^T$ o logarítmo da função de Verossimilhança é dado por:

$$L(\theta) = \sum_{i=1}^n \left[\log \left\{ \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \right\} + \phi \log \phi + y_i \log y_i - (\phi + y_i) \log(\mu_i + \phi) \right]$$

em que $\mu_i = \exp(x_i^T \beta)$.

- Estimativas de Máxima Verossimilhança, são obtidas com algoritmo iterativos como Newton-Raphson.
- Principal restrição do modelo é assumir relacionamento apenas linear das covariáveis com a resposta.
- Necessidade de expandir para efeitos espaciais e temporais.

Modelo aditivo generalizado (GAM)

- Proposto por Hastie e Tibishirani, 1990.
- É uma extensão do modelo linear generalizado.

$$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \dots$$

onde

$$\mu_i \equiv E(Y_i) \quad \text{e} \quad [Y_i] \sim \{\text{família exponencial}\}.$$

X_i^* é uma linha da matriz do modelo, para a parte estritamente paramétrica, θ é o vetor de parâmetros correspondentes e as f_j são funções suaves das covariáveis x_k .

- 1 Como representar as funções suaves.
- 2 Como estimar os parâmetros.

Representando uma função suave

- *Splines* Cúbicas é uma curva composta por seções de polinomiais cúbicas.
- Localização dos *knots* deve ser especificada.
- Escrever a base de *splines* cúbica.
- Suavidade do modelo é arbitrária.

Regressão Splines

- Uma alternativa para controlar o grau de suavização é adicionar uma penalidade na função objetivo:

$$\|y - X\beta\|^2 + \lambda \int_0^1 [f''(x)]^2 dx$$

$$\int_0^1 [f''(x)]^2 dx = \beta^T S \beta$$

onde S é uma matriz de coeficientes conhecidos.

Validação Cruzada Generalizada

- Se λ for muito pequeno alisar  muito os dados, se for grande n o alisar  de forma eficiente.
- A melhor escolha para o λ   aquela que mais aproxima a \hat{f} da real fun o f .
- Escore de valida o cruzada generalizada.

$$\nu_g = \frac{n \sum_{i=1}^n (y_i - \hat{f}_i)^2}{[\text{tr}(I - A)]^2}.$$

Resultados

Bairro	Obs.	Arma.	Início	Término	Sem.
Brasília Teimosa	2580	80	04/01/2005	15/05/2007	124
Casa Forte	1800	100	04/01/2006	16/05/2006	72
Dois Irmãos	1512	84	04/01/2005	16/05/2006	72
Engenho do Meio	3225	100	04/01/2005	15/05/2007	124
Morro Conceição	1800	100	04/01/2005	16/05/2006	72

Tabela: Descrição do experimento por bairros.

Resultados - Análise Fatorial

Tabela: Composição dos fatores de **Temperatura** em percentuais por bairro.

Semanas	BT	Fator	CFP	Fator	DI	Fator	MCP	fator	EM	Fator
Min 1	12.09	1	11.96	1	12.89	1	11.64	1	12.10	1
2	13.27	1	12.67	1	13.60	1	12.55	1	13.28	1
3	11.90	1	11.06	1	11.72	1	11.05	1	11.89	1
4	9.23	1	8.04	1	8.40	1	7.95	1	9.26	1
5	9.97	2	11.80	2	10.22	2	11.56	2	10.11	2
6	13.98	2	17.58	2	16.72	2	17.16	2	13.88	2
7	20.50	2	19.20	2	16.22	2	19.44	2	19.88	2
8	13.43	2	12.69	2	11.39	2	13.05	2	13.30	2
9	8.04	3	8.76	3	8.63	3	8.94	3	7.86	3
10	11.48	3	11.89	3	12.86	3	12.21	3	11.39	3
11	13.00	3	13.86	3	14.06	3	14.28	3	12.82	3
12	14.61	3	14.42	3	14.55	3	14.84	3	14.40	3
Max 1	15.02	1	14.06	1	15.68	1	14.28	1	14.94	1
2	15.18	1	13.67	1	14.63	1	13.86	1	15.24	1
3	12.87	1	11.50	1	12.45	1	11.61	1	12.95	1
4	10.43	1	9.60	1	10.63	1	9.75	1	10.33	1
5	12.32	2	7.45	1	11.07	2	7.30	1	12.41	2
6	16.00	2	13.99	2	12.26	2	14.18	2	16.20	2
7	13.79	2	13.91	2	12.02	2	13.74	2	14.21	2
8	6.27	3	10.83	2	10.11	2	10.87	2	6.26	3
9	8.98	3	9.89	3	9.83	3	9.47	3	9.12	3
10	11.53	3	12.07	3	11.79	3	11.85	3	11.66	3
11	13.29	3	14.43	3	14.00	3	14.15	3	13.53	3
12	12.79	3	14.67	3	14.28	3	14.26	3	12.97	3

Resultados - Análise Fatorial

Tabela: Composição dos fatores de **Umidade** em percentuais por bairro.

Semanas	BT	Fator	CFP	Fator	DI	fator	MCP	Fator
Precip 1	19.52	1	18.15	1	18.44	1	14.49	1
2	18.36	1	17.72	1	17.98	1	14.82	1
3	12.76	1	13.42	1	12.26	1	13.23	1
4	20.86	2	14.11	2	16.65	2	9.35	1
5	22.45	2	18.90	2	21.49	2	14.75	2
6	18.30	2	17.46	2	18.60	2	18.91	2
7	18.60	3	22.18	3	20.99	3	18.61	2
8	19.50	3	31.01	3	25.35	3	14.47	2
9	16.16	3	28.36	3	22.87	3	7.69	3
10	9.54	4	8.91	4	8.28	4	11.34	3
11	13.92	4	14.31	4	14.01	4	13.47	3
12	16.08	4	16.77	4	16.08	4	13.72	3
Umid 1	19.77	1	18.14	1	18.48	1	13.10	1
2	17.59	1	18.51	1	18.87	1	13.47	1
3	12.00	1	14.06	1	13.97	1	12.14	1
4	20.22	2	11.03	2	12.81	2	9.40	1
5	18.16	2	15.25	2	16.19	2	9.76	2
6	14.14	3	13.79	2	14.27	2	11.88	2
7	16.88	3	9.45	2	14.15	3	11.62	2
8	14.74	3	18.45	3	16.64	3	7.27	3
9	10.58	4	9.69	4	10.08	4	9.76	3
10	14.83	4	14.40	4	14.59	4	11.73	3
11	17.67	4	17.91	4	18.39	4	12.63	3
12	17.38	4	18.00	4	18.56	4	12.40	3

Aplicação do Modelo Estatístico

- 1 **Completo Covariáveis**
- 2 **Selecionado Covariáveis**
- 3 **Completo Climáticas**
- 4 **Selecionado Climáticas**
- 5 **Combinado**
- 6 **Espaço - temporal**

Aplicação do Modelo Estatístico

Tabela: Estimativa e p -valor do modelo combinado para todos os bairros.

Bairro	BT	CFP	DI	MCP	EM
Covariável	Resultado	Resultado	Resultado	Resultado	Resultado
Água Rede	Risco	NS	Risco	NS	NS
Canalizada	Risco	NS	NS	NS	NS
Abastecimento	NS	NS	Proteção	Proteção	NS
Quintal	NS	Risco	NS	NS	Proteção
Grande c	NS	NS	NS	NS	NS
Grande s	NS	NS	NS	NS	NS
Pequeno c	NS	NS	Proteção	Risco	Proteção
Pequeno s	NS	Proteção	Proteção	NS	NS
Fator risco	Risco	Risco	NS	NS	NS
Tipo Imóvel	NS	Risco	NS	NS	NS
Umidade 1	NS	Risco	NS	Risco	NA
Umidade 2	Risco	Risco	Risco	Risco	NA
Umidade 3	Risco	NS	NS	NS	NA
Umidade 4	NS	NS	NS	NS	NA
Temperatura 1	NS	NS	NS	NS	Proteção
Temperatura 2	NS	NS	NS	NS	NS
Temperatura 3	NS	Risco	Risco	Risco	Risco

Comparação via Critério de *Akaike*

Tabela: Comparação entre os modelos ajustados através do Critério de *Akaike*.

Modelos	Bairros				
	BT	CFP	DI	MCP	EM
Completo Covariáveis	37125	23398	19889	22838	43255
Selecionado Covariáveis	37114	23414	19883	22820	43244
Completo Climática	37032	23220	19730	22579	--
Selecionado Climáticas	37044	23215	19722	22579	--
Combinado	36996	23191	19626	22526	42691
Espaço-temporal	36772	23066	19379	22431	42540

Resultados - Dois Irmãos

Covariáveis	Estimativas	Erro Padrão	Valor t	p-valor
Intercepto	-1.926	0.662	-2.906	0.003
TEMP.3	0.213	0.024	8.830	$< 2e - 16$
UMIDADE.2	0.058	0.004	13.082	$< 2e - 16$
Pequeno s	0.407	0.103	3.933	$8.82e - 05$
Abastecimento	0.264	0.109	2.407	0.016
Suavização	<i>Edf</i>	Rank	F	p-valor
Coordenadas	23.48	29	10.30	$< 2e - 16$
Tempo	7.77	9	3.74	0.000

Tabela: Ajuste do Modelo Aditivo Generalizado para o bairro Dois Irmãos.

Resultados - Dois Irmãos

- Temperatura de 9 a 12 semanas.
- Umidade de 4 a 6 semanas.
- Recipientes pequenos sem tampa fator protetor - razão de ocorrência
50.3%(22.7% 84.2%)
- Frequência de abastecimento fator protetor - razão de ocorrência
30.2%(5.02% 61.5%)

Resultados - Dois Irmãos

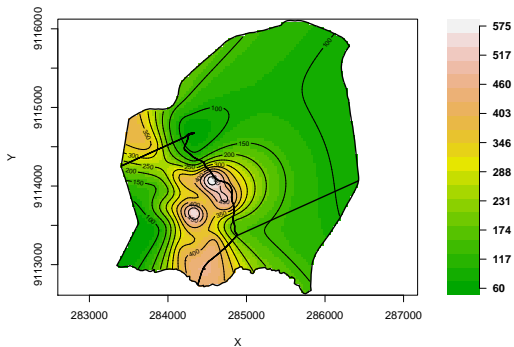


Figura: Superfície estimada da ocorrência de ovos de *Aedes aegypti*.

Resultados - Dois Irmãos

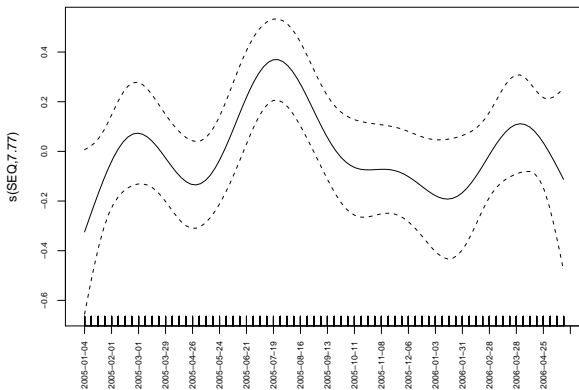


Figura: Ajuste do efeito temporal.

Resultados - Dois Irmãos

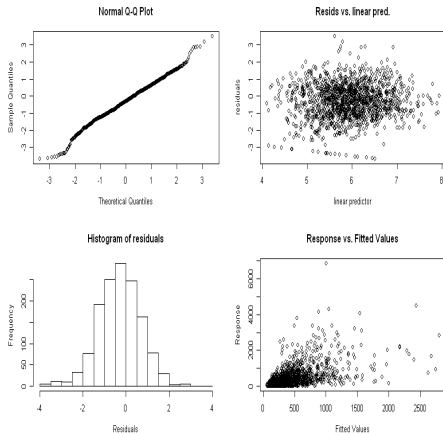


Figura: Análise dos resíduos.

Conclusões

- Processo complexo e delicado.
- Modelos estatísticos permitem uma visão clara dos fatores relevantes.
- Fatores climáticos (defasagens importantes).
- Efeito espacial.
- Efeito temporal.

Conclusões - Covariáveis da armadilha

- Priorizar as características:
 - 1 Plantas em vasos.
 - 2 Charco/poça.
 - 3 Garrafas e pneus.
 - 4 Piscinas e lajes sem cobertura.
- Preferência por residências.
- Abastecimento de água não constante.
- Recipientes pequenos sem tampa. Protetor ?
- Água ligada a rede geral / canaliza. Risco ?

Conclusões - Covariáveis climáticas

- Condições ideais
- 1 Período seco com altas temperaturas.(estocagem ?)
- 2 Período de \pm um mês para a reprodução do vetor.
- 3 Seguindo de um período de alta umidade.

Futuros trabalhos

- Ajuste de um modelo Binomial Negativa inflacionada de zeros.
- Discussão com profissionais da entomologia.
- Análise detalhada das observações de cada armadilha junto com a equipe de campo.
- Revisão dos resultados mediante diferentes formas de tratamentos para dados atípicos.
- Formas de predição espacial, para cenários futuros.
- Implementar os novos modelos no RDengue.
- Investigar a real necessidade do uso da análise fatorial.

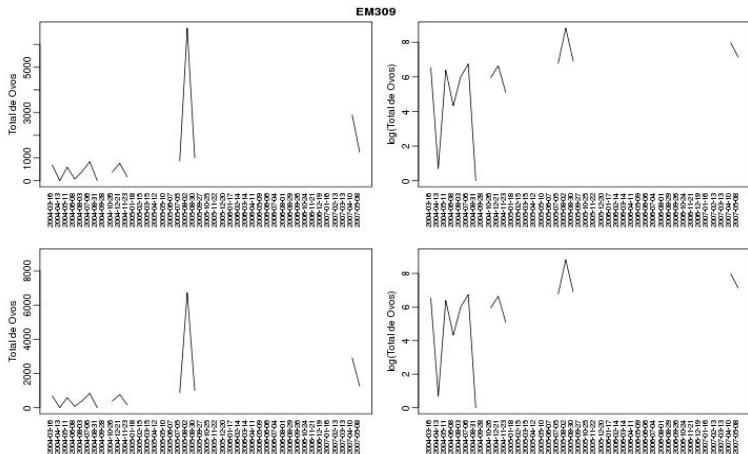
Limitações

- Experimento real, baixo nível de controle das causas de variação.
- Experimento *screening* ou exploratório.
- Confiabilidade dos dados.

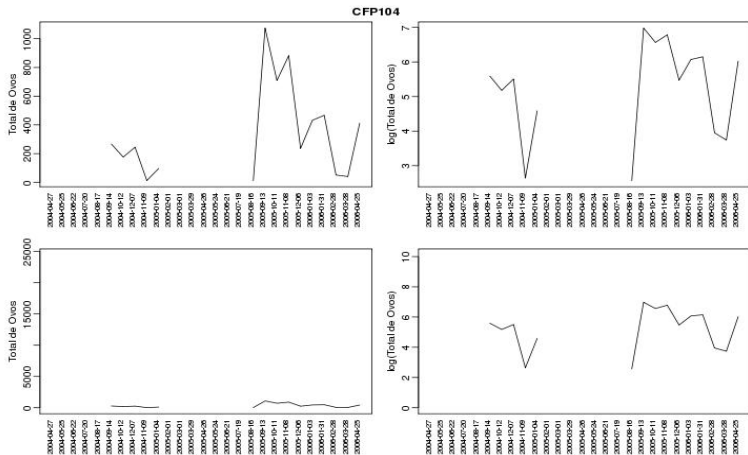
Um olhar crítico aos dados

- Covariáveis da armadilha.
- Quantidade excessiva de dados faltantes.
- Presença de picos inesperados.

Um olhar crítico aos dados



Um olhar crítico aos dados



Agradecimentos

- Equipes de campo e laboratório da Secretaria de Saúde de Recife/CVA e CPqAM/Fiocruz.
- Equipe do projeto SAUDAVEL.
- Dr. Antônio Miguel Vieira Monteiro.
- Dra. Lêda Régis, Wayner Vieira de Souza e a José Constantino Silveira.
- Professor PhD Paulo Justiniano Ribeiro Júnior.
- Laboratório de Estatística e Geoinformação (LEG-UFPR).