

# Modelagem Probabilística dos Casos de Óbitos por AIDS

**Danielle da Silva Pompeu**<sup>1 2</sup>

**Eudmar Paiva de Almeida**<sup>1</sup>

**Adrilayne dos Reis Araújo**<sup>3</sup>

**Terezinha Aparecida Guedes**<sup>4</sup>

## 1 Introdução

A Síndrome da Imuno deficiência Adquirida (AIDS) é causada pelo Vírus da Imuno deficiência Humana (HIV) e, é considerada pelo Ministério da Saúde no Brasil uma Doença Sexualmente Transmissível (DST), transmitida principalmente por meio da relação sexual, que é o seu principal meio de propagação, todavia, poderá também ocorrer através do sangue, assim como pelas secreções sexuais e pelo leite materno (BRASIL, 2011; UNICEF, 2011).

A epidemia do HIV/AIDS constitui-se em um fenômeno global cujo comportamento pandêmico tem afetado a humanidade independentemente da região geográfica, faixa etária, sexo, orientação sexual, raça, etc. (UNICEF, 2011).

Apesar dos primeiros casos de AIDS terem acontecidos em meados de 1977, nos Estados Unidos, somente em 1982, é que a doença foi definida com a Síndrome de Imuno Deficiência Adquirida. Segundo Shilts (1987), um dos primeiros casos descritos na literatura sobre a doença AIDS, foi com a morte da pesquisadora dinamarquesa, Margrethe P. Rask, no dia 12 de dezembro de 1977.

De acordo com Brasil (2012), os primeiros casos identificados de AIDS no Brasil ocorreram na década de 1980, sendo que os infectados eram predominantemente entre gays adultos, usuários de drogas injetáveis e hemofílicos. No período de 1980 a julho de 2011 foram notificados cerca de 608.230 casos de AIDS.

A AIDS segue aumentando em usuários de drogas e na transmissão sexual (Brasil, 2012). O controle e a prevenção são extremamente importantes, pois atualmente os métodos para que ocorra a prevenção são as mudanças no comportamento, principalmente para o uso de drogas e na relação sexual, como a utilização de preservativos. Neste sentido, este trabalho se justifica pelo fato de utilizar os métodos estatísticos capazes de proporcionar informações a respeito dos casos notificados de AIDS no Estado do Pará.

---

<sup>1</sup>PBE - UEM. e-mail: [danielle11silva@gmail.com](mailto:danielle11silva@gmail.com)

<sup>2</sup>Agradecimento ao CNPq pelo apoio financeiro.

<sup>1</sup>PBE - UEM. e-mail: [eudmar.almeida@gmail.com](mailto:eudmar.almeida@gmail.com)

<sup>3</sup>ICEN - UFPA. e-mail: [adrilaynereis@gmail.com](mailto:adrilaynereis@gmail.com)

<sup>4</sup>DES - UEM. e-mail: [taguedes@uem.br](mailto:taguedes@uem.br)

## 2 Material e métodos

### 2.1 Material

Os dados em estudos são referentes a 2.992 casos notificados de AIDS, no Estado do Pará, no período de janeiro de 2007 a julho de 2012. O banco de dados foi disponibilizado pelo Sistema de Informação de Agravos de Notificação (SINAN)/ Secretaria de Estado de Saúde Pública do Estado do Pará (SES-PA).

### 2.2 Métodos

Kutner *et al.* (2004) mostram que o objetivo da regressão logística é modelar uma variável resposta como função de uma ou mais variáveis preditoras que influenciam a sua ocorrência. Em que, a variável resposta é disposta em categorias e é expressa por meio de uma probabilidade de sucesso. Entretanto, mesmo quando a variável resposta não é disposta em categorias, pode-se dicotomizá-la de modo que a probabilidade de sucesso possa ser modelada a partir da Regressão Logística. Quando a variável resposta é disposta em duas ou mais categorias, a regressão logística é classificada em binária (aquela que aceita dois níveis de resposta), ordinal (aquela que segue uma ordenação natural) ou nominal (aquela que pode ter mais de três níveis sem uma ordenação natural).

Em um modelo logístico estima a probabilidade de determinada situação ocorrer ou não, com base em determinadas características e não a probabilidade de uma característica acontecer. Em um modelo de regressão linear simples (Kutner *et al.* 2004), dado por  $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ , se a variável resposta  $Y_i$  for binária, assumindo valores “0” ou “1” na ausência ou presença da característica em estudo, respectivamente, o modelo é dito ser de regressão logística. Em que,  $\varepsilon_i$  é o erro aleatório do modelo e  $X_i$  a variável independente.

Considerando apenas uma variável independente  $X_1$ , tem-se a Regressão Logística Simples. Neste caso, o modelo na sua forma usual é dado por

$$E(Y_{i,1}|X_{i,1}) = \frac{\exp(\beta_0 + \beta_1 X_i + \varepsilon_i)}{1 + \exp(\beta_0 + \beta_1 X_i + \varepsilon_i)},$$

em que,  $\beta_0$  e  $\beta_1$  são os coeficientes de regressão a serem estimados pelo método da máxima verossimilhança e  $X_i$  é a variável independente, onde  $i = 1, 2, \dots, n$ . No caso da Regressão Logística Múltipla, que é uma extensão do modelo logístico simples, o modelo é composto por duas ou mais variáveis independentes ( $X_1, X_2, \dots, X_n$ ) e, por seus, respectivos, coeficientes de regressão  $\beta_0, \beta_1, \dots, \beta_p$ . Portanto, tem-se  $\beta^t \mathbf{X} = \beta_0 + \beta_1 X_{i,1} + \dots + \beta_p X_{i,p}$ . Logo, o modelo se estende para o modelo logístico múltiplo, dado por  $\mathbf{Y}_i = E(\mathbf{Y}_i|\mathbf{X}_i) + \varepsilon_i$ , onde o termo  $\varepsilon_i$  é o erro aleatório do modelo e representa a diferença entre o valor observado de  $Y_i$  e o valor esperado condicionado de  $Y_i$  dado  $X_i$ , sendo  $Y_i$  uma variável dicotômica. Os valores dos parâmetros  $\beta_0, \beta_1, \dots, \beta_p$  são estimados a partir do método da máxima verossimilhança. De acordo com

Agresti (2003), uma das principais estatísticas utilizadas na análise de dados binários é a *razão de chances*, que é definida como a razão entre a chance de um evento ocorrer em um grupo e a chance de ocorrer em outro grupo, sendo que a *chance* é a probabilidade de ocorrência deste evento dividida pela probabilidade da não ocorrência do mesmo evento. A chance e a razão de chances são dadas respectivamente, por

$$chance_i = \frac{P(Y_i = 1)|X_i}{P(Y_i = 0)|X_i} = \frac{\pi(X_i)}{1 - \pi(X_i)} \quad e \quad RC = \frac{\frac{\pi(X_i=1)}{1-\pi(X_i=1)}}{\frac{\pi(X_i=0)}{1-\pi(X_i=0)}}$$

### 3 Resultados e discussões

#### 3.1 Aplicação da regressão logística

Para o processo de modelagem a partir da regressão logística binária, a variável resposta evolução da doença ( $Y_i$ ) foi codificada em 1, se o paciente evoluiu a óbito e 0, se o paciente está vivo. Considerando as variáveis predictoras, Gênero que foi codificada como 1, se o paciente é do gênero Masculino e 0, se o paciente é do gênero feminino e, finalmente, a variável preditora uso de drogas foi codificada como 1, se o paciente utiliza droga e 0, se o paciente não utiliza droga. Dessa maneira, o modelo de Regressão Logística Múltipla ajustado é dado por

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 Genero_i + \beta_2 Uso\ de\ Droga_i)}{1 + \exp(\beta_0 + \beta_1 Genero_i + \beta_2 Uso\ de\ Droga_i)} \quad (1)$$

sendo que as variáveis inseridas no Modelo (1) foram selecionadas a partir do Método *Stepwise*, as quais foram mantidas no modelo ajustado apenas as variáveis significantes, considerando a probabilidade de entrada de 20% e a de saída de 5% a cada passo do *Stepwise*. A Tabela 1 mostra as estimativas dos parâmetros do modelo ajustado, seus erros padrão, os valores do nível descritivo  $p$ , as razões de chances, bem como o intervalo de confiança de 95%, para a probabilidade de ocorrência de óbito por AIDS no Estado do Pará, no período de janeiro de 2007 a julho de 2010. Na Tabela 1, observa-se que o coeficiente positivo ( $\hat{\beta}_1 = 0,99299$ ) para a variável gênero sugere que pacientes do gênero masculino apresentam maior probabilidade de incidência de óbito por AIDS em relação aos pacientes do gênero feminino, ou seja, paciente do gênero masculino têm aproximadamente 3 vezes mais chance de evoluir á óbito por AIDS do que paciente do gênero feminino, desde que as demais variáveis permaneçam constante. O coeficiente positivo ( $\hat{\beta}_2 = 1,25149$ ) para a variável uso de droga, explica que paciente que utilizam drogas tem maior probabilidade de incidência de óbito por AIDS em relação aos pacientes que não utilizam drogas, ou seja, os usuários de drogas tem aproximadamente 4 vezes mais chance de evoluir á óbito por AIDS do que paciente que não utilizam, permanecendo as demais variáveis constante.

Tabela 1: Estatísticas resultantes da Regressão Logística Binária no processo de modelagem da probabilidade de ocorrência de óbito por AIDS, no Estado do Pará, no período de janeiro de 2007 a julho de 2012.

Preditor	Estimativa	Erro Padrão	Z	p	RC	IC(95%)-LI	IC(95%)-LS
Constante	-2,62454	0,175792	-14,93	0,000	-	-	-
Gênero	0,99299	0,191511	5,19	0,000	2,7	1,85	3,93
Uso de Droga	1,25149	0,258428	4,84	0,000	3,5	2,11	5,80

A Tabela 2 mostra os testes qui-quadrado de bondade de ajuste para os métodos de Pearson, Deviance e Hosmer Lemeshow que verificam a hipótese  $H_0$ : o ajuste dos dados é bom versus  $H_1$ : o ajuste dos dados não é bom. Pode-se observar que, ao nível de significância de 5%, há evidências suficientes para afirmar que o ajuste dos dados é bom, pois os valores do p (nível descritivo), para cada método, são maiores que 0,05.

Tabela 2: Testes Qui-quadrado de bondade de ajuste pelos métodos Pearson, Deviance e Hosmer lemeshow para a probabilidade de ocorrência de óbito por AIDS, no Estado do Pará, no período de janeiro de 2007 a julho de 2012.

Método	Qui-quadrado	g.l	P
Person	2,48686	1	0,115
Deviance	2,10241	1	0,147
Hosmer lemeshow	0,09585	1	0,757

A Tabela 3 apresenta as probabilidades ajustadas para o modelo de regressão logística múltipla ajustado, para a probabilidade de óbito por AIDS, no estado do Pará, no período de janeiro de 2007 a julho de 2012. Observa-se por meio desta que fixando a variável gênero, os pacientes do sexo masculino que utiliza drogas têm maior probabilidade de vir a óbito por AIDS, e um paciente do gênero feminino que não utiliza droga tem menor probabilidade de vir a óbito por AIDS.

Tabela 3: Probabilidades ajustadas a partir do modelo de Regressão Logística Múltipla Ajustado, para a probabilidade de ocorrência de óbito por AIDS, no Estado do Pará, no período de janeiro de 2007 a julho de 2012.

Gênero	Uso de Droga	Probabilidade (%)
Masculino	Não usa	16,36%
Masculino	Usa	40,61%
Feminino	Não usa	6,76%
Feminino	Usa	20,21%

## 4 Conclusões

Este trabalho teve como objetivo desenvolver um modelo logístico capaz de mostrar a probabilidade de ocorrência de óbito por AIDS, no estado do Pará, no período de janeiro de 2007 a julho de 2012. Observou-se que os testes de bondade de ajuste mostram que o modelo obtido se ajusta bem aos dados. A partir da regressão logística pode-se observar que, os pacientes do gênero masculino que usam drogas têm 40,61% de chance de vir a óbito por AIDS, em relação aos pacientes que não usam drogas, pode-se observar que os pacientes do gênero feminino que usa drogas têm 20,21% mais chance de vir a óbito por AIDS em relação aos pacientes que não usam drogas. Já os pacientes que usam drogas têm 50% de chance a mais de vir a óbito por AIDS, sendo mantida a mesma característica a variável gênero.

## Referências

- [1] AGRESTI, A. Categorical data analysis, Second Edition, Jonh Wileyand Sons: New York. 2003.
- [2] BRASIL. Ministério da saúde. Secretaria de vigilância em saúde. Departamento de DST e AIDS. Boletim epidemiológico AIDS/DST, Brasília: Ministério da Saúde, 2012.
- [3] BRASIL. Ministério da saúde. Secretaria de vigilância em saúde. Programa nacional de DST, AIDS e hepatites. Boletim epidemiológico AIDS/DST. Brasília: Ministério da Saúde, 2011.
- [4] KUTNER, M. H.; NETER, J.; NACHTSHEIM, C. J. et al. Applied linear statistical models. 5<sup>th</sup>.ed., Illinois: Burr Ridger, 2004.
- [5] SHILTS, Randy. O prazer com risco de vida. Record. Rio de Janeiro, 1987.
- [6] UNICEF. Fundo das Nações Unidas para a Infância. Situação da Adolescência Brasileira 2011 - O Direito de Ser Adolescente: Oportunidade para reduzir vulnerabilidades e superar desigualdade. Brasília: UNICEF, 2011.