

Modelos para Dados de Contagem

Katiane S. Conceição

SME/ICMC/USP

Outubro/2017

- 1 Introdução
- 2 Motivação
- 3 Aplicações: Dados Reais
- 4 Modelo de Regressão
- 5 Trabalhos em Desenvolvimento

Alguns Conceitos

- **Estatística:** É uma ciência baseada na **Teoria da Probabilidade**, cujo o objetivo principal é nos auxiliar a tomar decisões ou tirar conclusões em situações de incerteza, a partir dos dados.

Observação: A coleta de dados representa somente um dos aspectos da Estatística.

Alguns Conceitos

- **Estatística:** É uma ciência baseada na **Teoria da Probabilidade**, cujo o objetivo principal é nos auxiliar a tomar decisões ou tirar conclusões em situações de incerteza, a partir dos dados.

Observação: A coleta de dados representa somente um dos aspectos da Estatística.

Resumo: Estatística \Leftrightarrow Incerteza

Exemplos:

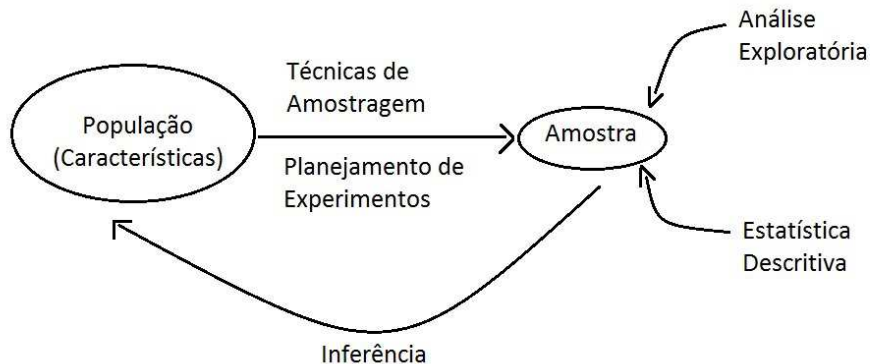
- Experimentar a quantidade de sal de um determinado prato;
- Incluir na receita de bolo uma xícara de farinha de trigo;
- Olhar para o céu e decidir se levará ou não o guarda-chuva;
- Listar os possíveis números que ocorrerão na próxima aposta da mega-sena;
- Proferir o possível ganhador em um confronto de times de futebol.

Alguns Conceitos

- **Variável:** Qualquer característica de interesse associada aos elementos de uma população.
- **Variável Aleatória:** É uma função que X que associa a cada elemento do espaço amostral a um número real.

$$X : \Omega \longrightarrow R$$

Ideia



Exemplos Pátricos

Exemplo 1:

- ε : Descrever para os confrontos de um determinado time uma das classificações: “ganhou” ou “não ganhou”;
- Ω : {ganhou; não ganhou};
- $X = \begin{cases} 0, & \text{se o time não ganhou} \\ 1, & \text{se o time ganhou} \end{cases}$.
- $\mathcal{R}_x = \{0, 1\}$

Exemplos Pátricos

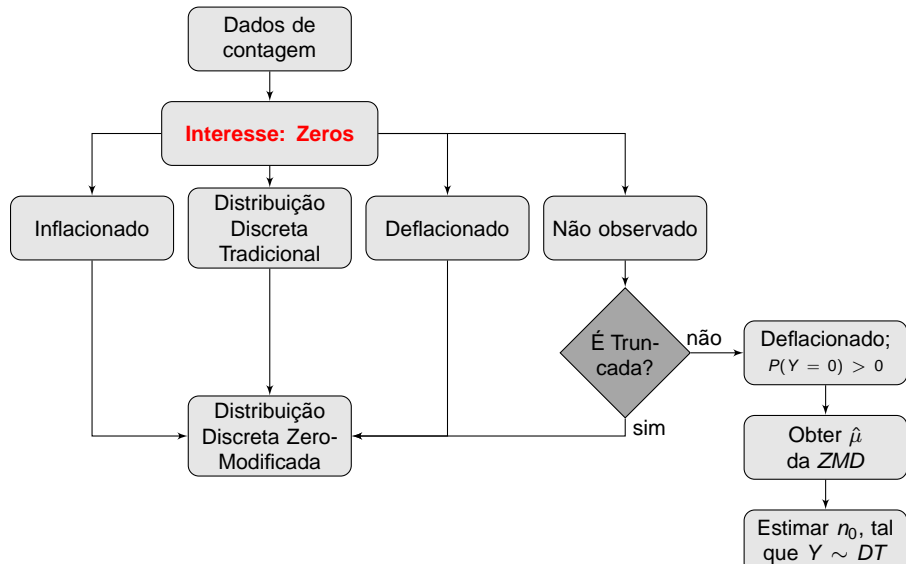
Exemplo 2:

- ε : Observar o número de gols marcados por um determinado time em uma partida;
- $\Omega : \{0, 1, 2, \dots\}$;
- $X =$ número de gols marcados por um determinado time em uma partida;
- $\mathcal{R}_x = \{0, 1, 2, \dots\}$

Exemplos Pátricos

Exemplo 3:

- ε : Notificação de uma determinada doença;
- $\Omega : \{0, 1, 2, \dots\}$;
- X = número de notificações de uma determinada doença em uma cidade;
- $\mathcal{R}_X = \{0, 1, 2, \dots\}$



Objetivos Específicos

- Com a família de distribuições zero-modificadas proposta, conseguimos:
 - Maior flexibilidade para analisar dados de contagem;
 - Não exige o conhecimento prévio da frequência de zeros na amostra;
 - Responder adequadamente as chances de ocorrências dos eventos;
 - Consequentemente, tirar conclusões fidedignas.

Exemplo: Número de derrotas até a vitória

- Histórico dos confrontos entre Real Madrid CF and FC Barcelona (Março 1916 a Abril 2014).
- 179 jogos: 91 vitórias do Real Madrid CF e 88 vitórias do FC Barcelona.
- **Dados:** Número de vitórias do FC Barcelona até a ocorrência de uma vitória do Real Madrid CF.

Tabela: Distribuição de frequência do número de vitórias do FC Barcelona até a ocorrência de uma vitória do Real Madrid CF between Março/1916 e Abril/2014.

y_i	0	1	2	3	4	5
f_i	39	31	11	6	3	1

Exemplo: Número de derrotas até a vitória

Questões:

- Qual a probabilidade do Real Madrid CF ganhar do FC Barcelona logo após uma vitória?
- Qual a probabilidade do Real Madrid CF ganhar do FC Barcelona exatamente após 4 partidas?

Exemplo: Número de gols marcados

Tabela: Distribuição de frequência do número de gols marcados por partida de alguns times do Brasileirão Série A em 2016.

Classificação	Times	Gols Marcados					
		0	1	2	3	4	5
1	Palmeiras	6	13	13	3	3	-
2	Santos	9	9	11	8	1	-
3	Flamengo	8	11	16	3	-	-
4	Atlético MG	7	13	9	7	1	1
7	Corinthians	13	10	10	3	2	-
10	São Paulo	13	13	8	2	1	1
14	Sport	12	14	7	1	2	2
17	Internacional	14	14	9	1	-	-
18	Figueirense	16	16	4	2	-	-
19	Santa Cruz	15	11	6	3	2	1
20	América MG	18	17	3	-	-	-

Exemplo: Número de gols marcados

Questões:

- Há uma probabilidade alta (inflação), satisfatória (tradicional) ou baixa (deflação) de um determinado time não marcar gol?
- Qual a probabilidade de um determinado time marcar k gols?
- Qual a probabilidade de um determinado time ganhar em um próximo confronto?

Exemplo: Notificações de leptospirose

Tabela: Distribuição de frequência de notificações de leptospirose nas cidades do Estado da Bahia em 2004.

y_i	0	1	2	3	4	5	6	11	12	14	16	111
f_i	365	29	7	3	2	2	1	2	1	1	1	1

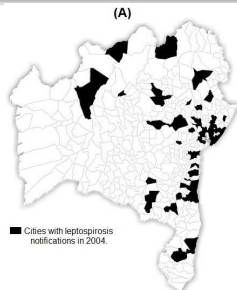


Figura: Mapa do Estado da Bahia: cidades que apresentaram notificações de leptospiros em 2004.

Modelo de Regressão

- Muitas vezes há interesse em explicar o conjunto de dados de contagem utilizando-se variáveis exógenas (ou explicativas);
- O procedimento para isso é considerar modelos de regressão.

Exemplos:

- Avaliar a número de gols marcado em função do investimento feito no time.
- Avaliar a notificação de uma determinada doença em função do IDH da localidade.

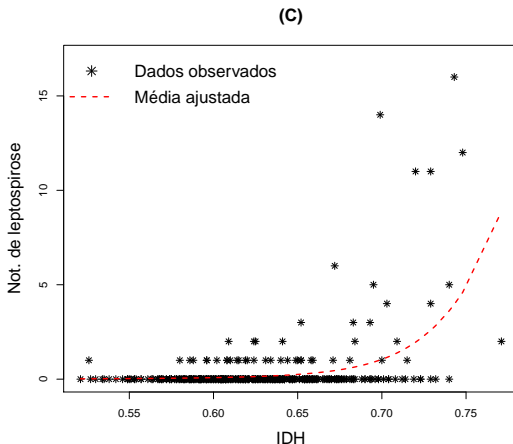


Figura: Média ajustada em função do IDH.

Propostas

- Estender a ideia da modificação para uma observação k :
 - Pandey(1965) investigou o número de flores produzidos por plantas *Primula veris*, notando um número excessivo de plantas com 8 flores.
 - número de filhos em famílias de uma localidade que adota políticas para incentivar o aumento da taxa de fecundidade em forma de subsídios para os pais(deflação na observação $k = 1$, um único filho), evitando uma diminuição maior da população (exemplos: Alemanha, Austrália, Estônia, Escócia e Suécia).
 - a quantidade de um determinado produto comprado por cliente em uma rede de supermercados, sendo que há uma oferta que propõe uma diminuição no preço da unidade deste produto caso o cliente opte por levar acima de k unidades ($k > 1$) (deflação da observação k , de itens vendidos).