

Modelagem de fração de cura, aplicado na classificação de clientes com dados segmentados

Douglas de Paula Nestlehner
Orientador: José Carlos Fogo

Universidade Federal de São Carlos

20 de Março de 2023

① Introdução

② Metodologia

- Modelo de Mistura Padrão
- Modelo de Mistura Padrão com Regressão

③ Aplicação

- Análise Exploratória
- Ajuste dos Modelos
- Comparação

④ Conclusão

Introdução

- Estudos de fidelização de clientes de serviço por assinatura:

- Estudos de fidelização de clientes de serviço por assinatura:
 - Efeitos econômicos da pandemia;

- Estudos de fidelização de clientes de serviço por assinatura:
 - Efeitos econômicos da pandemia;
 - Alta concorrência;

- Estudos de fidelização de clientes de serviço por assinatura:
 - Efeitos econômicos da pandemia;
 - Alta concorrência;
 - Novas tecnologias.

- Estudos de fidelização de clientes de serviço por assinatura:
 - Efeitos econômicos da pandemia;
 - Alta concorrência;
 - Novas tecnologias.
- Interesse em estudar a ocorrência de *churn*.

- Estudos de fidelização de clientes de serviço por assinatura:
 - Efeitos econômicos da pandemia;
 - Alta concorrência;
 - Novas tecnologias.
- Interesse em estudar a ocorrência de *churn*.
- Dados com características:

- Estudos de fidelização de clientes de serviço por assinatura:
 - Efeitos econômicos da pandemia;
 - Alta concorrência;
 - Novas tecnologias.
- Interesse em estudar a ocorrência de *churn*.
- Dados com características:
 - Alto numero de não ocorrência do evento de interesse, o churn;

- Estudos de fidelização de clientes de serviço por assinatura:
 - Efeitos econômicos da pandemia;
 - Alta concorrência;
 - Novas tecnologias.
- Interesse em estudar a ocorrência de *churn*.
- Dados com características:
 - Alto número de não ocorrência do evento de interesse, o churn;
 - Segmentação por diversas características.

- Estudos de fidelização de clientes de serviço por assinatura:
 - Efeitos econômicos da pandemia;
 - Alta concorrência;
 - Novas tecnologias.
- Interesse em estudar a ocorrência de *churn*.
- Dados com características:
 - Alto numero de não ocorrência do evento de interesse, o churn;
 - Segmentação por diversas características.
- Probabilidade do cliente cancelar o serviço ?

Objetivo

Utilizar modelos de regressão com fração de cura no intuito de prever a probabilidade de um indivíduo permanecer fiel por um longo período, considerando suas características.

Utilizar modelos de regressão com fração de cura no intuito de prever a probabilidade de um indivíduo permanecer fiel por um longo período, considerando suas características.

- Modelo de mistura padrão;

Utilizar modelos de regressão com fração de cura no intuito de prever a probabilidade de um indivíduo permanecer fiel por um longo período, considerando suas características.

- Modelo de mistura padrão;
- Modelo de mistura padrão com Regressão.

Metodologia

Modelo de Mistura Padrão

- Proposto por Berkson & Gage (1952);

Modelo de Mistura Padrão

- Proposto por Berkson & Gage (1952);
- Consiste na mistura de duas distribuições paramétricas (função de sobrevivência própria e imprópria):

Modelo de Mistura Padrão

- Proposto por Berkson & Gage (1952);
- Consiste na mistura de duas distribuições paramétricas (função de sobrevivência própria e imprópria):

$$S_{pop}(t_i) = p + (1 - p)S(t_i),$$

$$f_{pop}(t_i) = (1 - p)f(t_i),$$

$$L(\theta|D) = \prod_{i=1}^n [f_{pop}(t_i)]^{\delta_i} [S_{pop}(t_i)]^{1-\delta_i}$$

Modelo de Mistura Padrão Weibull

- Abordamos as distribuições Exponencial, Weibull e Gompertz;

Modelo de Mistura Padrão Weibull

- Abordamos as distribuições Exponencial, Weibull e Gompertz;
- Caracterizando o tempo até a ocorrência do evento de interesse com a distribuição Weibull, obtemos:

Modelo de Mistura Padrão Weibull

- Abordamos as distribuições Exponencial, Weibull e Gompertz;
- Caracterizando o tempo até a ocorrência do evento de interesse com a distribuição Weibull, obtemos:

$$S_{pop_k}(t_{ki}) = p_k + (1 - p_k)e^{-(t_{ki}\lambda_k)^{\alpha_k}},$$

e

$$f_{pop_k}(t_{ki}) = (1 - p_k)\alpha_k\lambda_k(t_{ki}\lambda_k)^{(\alpha_k-1)}e^{-(t_{ki}\lambda_k)^{\alpha_k}}$$

Modelo de Mistura Padrão Weibull

- E conseqüentemente, obtemos a função de verossimilhança do k -ésimo grupo:

$$L_k(\theta_k|D) = \prod_{i=1}^{n_k} \left[(1 - p_k) \alpha_k \lambda_k (t_{ki} \lambda_k)^{(\alpha_k - 1)} e^{-(t_{ki} \lambda_k)^{\alpha_k}} \right]^{\delta_{ki}} \left[p_k + (1 - p_k) e^{-(t_{ki} \lambda_k)^{\alpha_k}} \right]^{1 - \delta_{ki}}$$

Modelo de Mistura Padrão Weibull com Regressão

- Considerando o efeito na fração de curados e não curados, Kuk e Chen (1978) definem:

$$S_{pop}(t|Z, X) = p(X) + (1 - p(X))S(t)^{\exp(Zb)},$$

e

$$f_{pop}(t|Z, X) = (1 - p(X))f(t)\exp(Zb)S(t)^{\exp(Zb)-1}.$$

Modelo de Mistura Padrão Weibull com Regressão

- Em que, a fração de cura $p(X)$ pode ser modelada por:

$$p(X) = \frac{\exp(\beta X)}{1 + \exp(\beta X)}, \quad (\text{Ligação logito})$$

$$p(X) = \Phi(\beta X), \quad (\text{Ligação probito})$$

$$p(X) = \exp(-\exp(\beta X)). \quad (\text{Ligação complemento log-log})$$

Aplicação

Aplicação dos estudos em uma base de dados de planos de internet, disponível na plataforma [Maven Analytics](#).

Aplicação dos estudos em uma base de dados de planos de internet, disponível na plataforma [Maven Analytics](#).

- Dados coletados no segundo trimestre de 2022;

Aplicação dos estudos em uma base de dados de planos de internet, disponível na plataforma [Maven Analytics](#).

- Dados coletados no segundo trimestre de 2022;
- Obter informações sobre *churn* de seus clientes.

Aplicação dos estudos em uma base de dados de planos de internet, disponível na plataforma [Maven Analytics](#).

- Dados coletados no segundo trimestre de 2022;
- Obter informações sobre *churn* de seus clientes.
- Informações dos clientes que adquiriram o produto da empresa nos últimos 72 meses da data de coleta.

A base completa traz informações dos últimos 72 meses, totalizando 7043 clientes, em que cada registro de cliente contém detalhes sobre seus dados de assinatura.

Tabela 1: Representação da base de dados.

ID	Tipo de Internet	Status	Tempo
0002-ORFBO	Cable	Stayed	58
0003-MKNFE	Fiber Optic	Churned	9
0004-TLHLJ	DSL	Churned	39
0011-IGKFF	Plus	Stayed	13
0013-EXCHZ	Cable	Churned	71
⋮	⋮	⋮	⋮

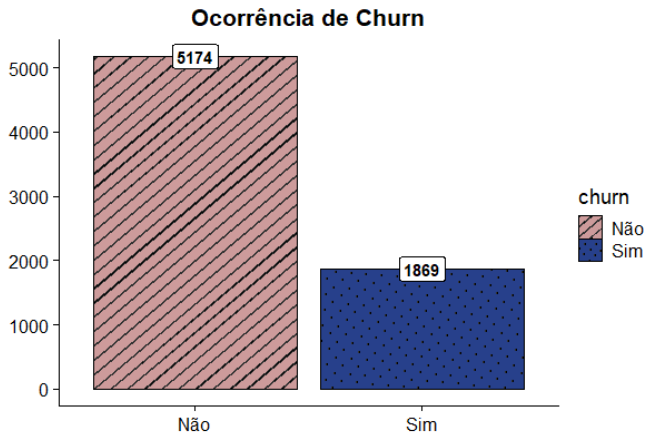


Figura 1: Gráfico de barras da quantidade de ocorrências ou não de *churn*.

Tabela 2: Ocorrência de *churn* em cada plano.

	Cable	Fiber Optic	DSL	Plus
Ocorrência <i>churn</i>	25.66%	40.25%	18.58%	7.40%
Censura	74.34%	59.75%	81.42%	92.60%

Proporção de Churn por plano

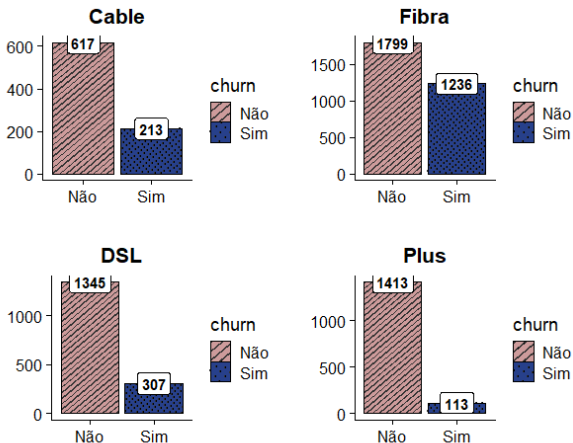


Figura 2: Gráfico de barras da quantidade de ocorrências ou não de *churn* segmentado por tipo de plano.

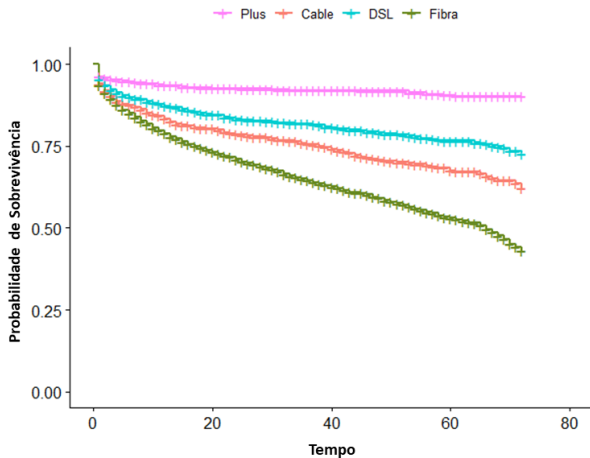


Figura 3: Curvas de sobrevivência estimadas pelo método de Kaplan-Meier para cada grupo.

Ajustes

Modelo de Mistura padrão Weibull

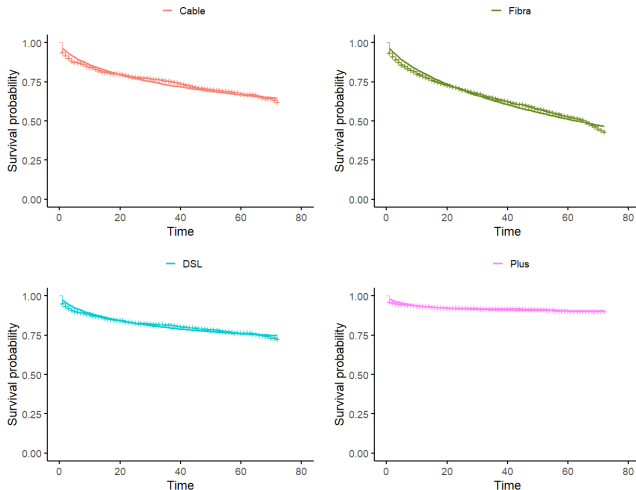


Figura 4: Estimativas obtida no modelo de mistura padrão Weibull.

Tabela 3: Parâmetros estimados do modelo de mistura padrão Weibull

Parâmetro	Segmentação	Estimativa	IC (95%)
p	Cable	0.6326	(0.5816 - 0.6808)
p	Fibra	0.3855	(0.3449 - 0.4277)
p	DSL	0.7440	(0.7143 - 0.7715)
p	Plus	0.9051	(0.8865 - 0.9210)
λ	Cable	51.801	(19.917 - 134.72)
λ	Fibra	106.00	(95.400 - 117.00)
λ	DSL	32.407	(18.956 - 55.401)
λ	Plus	9.9599	(6.4968 - 15.269)
α	Cable	0.6721	(0.5731 - 0.7882)
α	Fibra	0.7100	(0.6750 - 0.7460)
α	DSL	0.6905	(0.6109 - 0.7805)
α	Plus	0.6800	(0.5802 - 0.7970)

- Realizamos o mesmo procedimento para as distribuições Exponencial e Gompertz;
- Pelos critérios de AIC e BIC, o modelo considerando a Weibull apresentou melhores resultados

Modelo de Mistura Padrão Weibull com Regressão

Modelo de Mistura Padrão Weibull com Regressão

- Consideramos inicialmente a covariável “Idade”;

Modelo de Mistura Padrão Weibull com Regressão

- Consideramos inicialmente a covariável “Idade”;
- Ajustamos um modelo para cada segmentação.

Modelo de Mistura Padrão Weibull com Regressão

- Consideramos inicialmente a covariável “Idade”;
- Ajustamos um modelo para cada segmentação.

Tabela 4: Parâmetros estimados do modelo de regressão Weibull na presença da covariável idade, para a segmentação Cable.

Parâmetro	Estimativa	Desvio
λ	-2.1363	0.4116
α	-0.3549	0.0744
β_0	1.3470	0.4290
β_1	-0.0243	0.0087
b_1	-0.0087	0.0074

$$p_{(\text{cable})}(X) = \frac{\exp(\beta X)}{1 + \exp(\beta X)}$$

$$p_{(\text{cable})}(X) = \frac{\exp(\beta_0 + \beta_1 \cdot \text{Idade})}{1 + \exp(\beta_0 + \beta_1 \cdot \text{Idade})}$$

$$p_{(\text{cable})}(X) = \frac{\exp(1.3470 - 0.0243 \cdot 44.70)}{1 + \exp(1.3470 - 0.0243 \cdot 44.70)}$$

$$p_{(\text{cable})}(X) = 0.5644$$

- Ou seja, a fração de cura estimada pelo modelo de regressão considerando a covariável idade, foi de 56.44%.

$$\begin{aligned} p_{(\text{fibra})}(X) &= \frac{\exp(0.1056 - 0.0115 \cdot 49.80)}{1 + \exp(0.1056 - 0.0115 \cdot 49.80)} \\ &= 0.3851 \end{aligned}$$

$$\begin{aligned} p_{(\text{DSL})}(X) &= \frac{\exp(2.1242 - 0.0276 \cdot 44.83)}{1 + \exp(2.1242 - 0.0276 \cdot 44.83))} \\ &= 0.7086 \end{aligned}$$

$$\begin{aligned} p_{(\text{plus})}(X) &= \frac{\exp(2.2412 - 0.0002 \cdot 42.77)}{1 + \exp(2.2412 - 0.0002 \cdot 42.77)} \\ &= 0.9032 \end{aligned}$$

Tabela 5: Comparação entre os modelos ajustados.

Segmentação	Modelo	p	Censuras	AIC	BIC
Cable	MW	0.6326	0.7434	2393.941	2408.105
	MWCov	0.5644		2389.385	2412.993
Fibra	MW	0.3855	0.5975	13140.490	13158.550
	MWCov	0.3851		13258.010	13282.090
DSL	MW	0.7440	0.8142	3618.945	3635.174
	MWCov	0.7086		3609.125	3636.173
Plus	MW	0.9051	0.9260	1430.704	1446.695
	MWCov	0.9032		1434.281	1460.933

Conclusão

- Estudo dos modelos de mistura padrão;
- Aplicação em um problema de fidelização de clientes de serviços por assinatura;
- Segmentação do estudo, permitindo observar a aplicação em diversas situações;
- Obtendo resultados relevantes para a proposta da aplicação.

Obrigado!