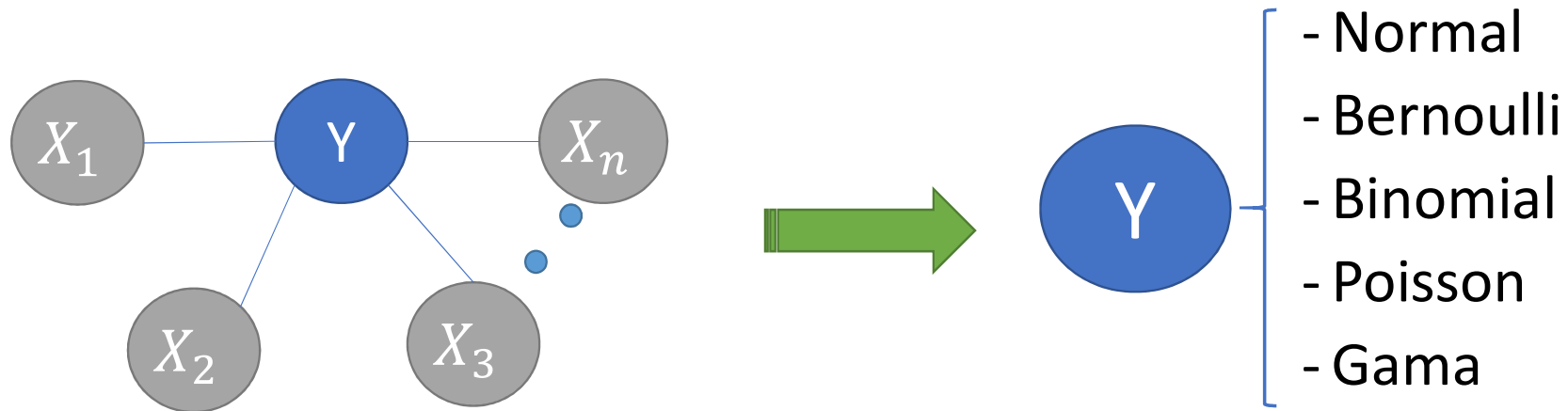




PUC Minas
Virtual

Síntese do Método GLM

Fenômeno a ser estudado



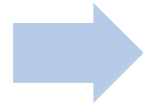
Diagnóstico sobre Y

- Fundamental importância
- Elaborado antes da criação dos modelos
- Auxilia na escolha do melhor modelo

COMPONENTES DE UM GLM

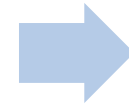
1° Componente

- Componente aleatório. Variável resposta (Y)
- segue uma distribuição que pertence à família exponencial.



2° Componente

- Componente sistemático
- É a parte fixa do modelo - as covariáveis (estrutura linear do nosso modelo)



3° Componente

- vai ligar o componente aleatório com o componente sistemático.
- Chamaremos de função de ligação(*link function*)

Identificando os componentes em um modelo clássico

$$Y = \mu + \varepsilon$$

Onde:

- **Y**: vetor da variável resposta
- **$\mu = E(Y) = X\beta$** , o componente sistemático
- **X**: Matriz de Dimensões
- **β** : Vetor de coeficientes
- **E**: Componente aleatório
- $Y \sim N(\mu, \sigma^2)$;

- Identificando a função de ligação : $g(\mu) = XB$
- No caso do modelo clássico a função de $g(\mu) = \mu$ (identidade)

DEFININDO O GLM ALGEBRICAMENTE

$$\eta = \alpha + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_n X_{ni} + e_i$$

Onde:

η : função de ligação canônica

α : representa a constante

$\hat{\beta}_1$ a $\hat{\beta}_k$: coeficientes de cada uma das variáveis independentes $X_{1i}, X_{2i}, \dots, X_{ni}$.

GLM'S E SUAS CARACTERÍSTICAS

Modelo	Característica de Y	Distribuição	Função de ligação Canônica (η)
Linear	Quantitativa	Normal	\hat{Y}
Logística binária	Qualitativa com 2 categorias (Dummy)	Bernoulli	$\ln\left(\frac{p}{1-p}\right)$
Logística Multinomial	Qualitativa com 3 ou mais categorias	Binomial	$\ln\left(\frac{p_m}{1-p_m}\right)$
Poisson	Quantitativa com dados de contagem	Poisson	$\ln(\lambda)$
Binomial Negativa	Quantitativa com dados de contagem	Poisson-gama	$\ln(\lambda)$

Especificações dos modelos lineares generalizados

Linear:

$$\hat{Y} = \alpha + \hat{\beta}_1 X_{1i} + \hat{\beta}_1 X_{2i} + \dots + \hat{\beta}_n X_{nn} + e_i$$

**Logística
Binária:**

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \hat{\beta}_1 X_{1i} + \hat{\beta}_1 X_{2i} + \dots + \hat{\beta}_n X_{nn}$$

**Logística
Multinomial:**

$$\ln\left(\frac{p_m}{1-p_m}\right) = \alpha + \hat{\beta}_1 X_{1i} + \hat{\beta}_1 X_{2i} + \dots + \hat{\beta}_n X_{nn}$$

**Dados de
Contagem:**

$$\ln(\lambda) = \alpha + \hat{\beta}_1 X_{1i} + \hat{\beta}_1 X_{2i} + \dots + \hat{\beta}_n X_{nn}$$

Pressupostos da Regressão

A variável resposta precisa ter as características da família exponencial

As preditoras podem ser quantitativas ou categóricas

Ausência de autocorrelação

Relação linear entre o vetor das variáveis explicativas X e a variável independente Y ;

Ausência de correlação entre os resíduos e as variáveis explicativas (Multicolinearidade)

Assume que as observações são independentes

Análise de Diagnóstico

Análise de Diagnósticos

Análise de
Resíduos

- Analisar algumas suposições.
- Verificar se o ajuste dos dados está correto (adequado)

Métodos de
diagnóstico

avaliar pontos
que possam ser
influentes

Diagnóstico do Ajuste

A análise de Diagnóstico é uma etapa fundamental no ajuste dos modelos de regressão.

Como garantimos que o modelo está ajustado aos dados?

Seu principal objetivo é a avaliação do modelo ajustado. No caso dos GLMs, iremos verificar os seguintes itens:

Distribuição
proposta

Parte
sistemática
proposta

Função de
ligação

Efeito de
observações
mal
ajustadas

Identificação
de pontos
influentes

Análise dos resíduos

Resíduos de Pearson

Resíduos de Deviance

Gráfico de Resíduos



PUC Minas
Virtual