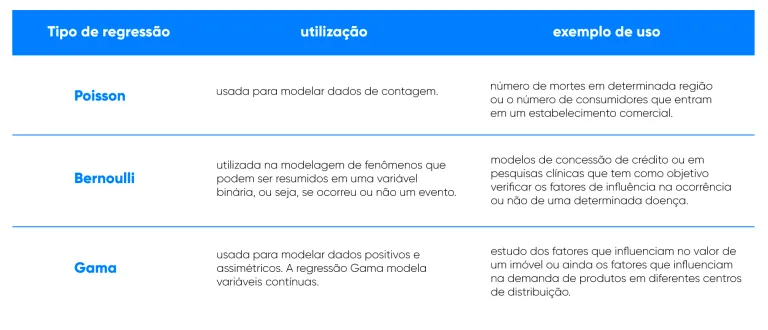
O que um MLG precisa ter?

<https://operdata.com.br/blog/o-que-sao-modelos-lineares-generalizados/>



Os MLGs (Modelos Lineares Generalizados) são uma extensão dos modelos de regressão simples e múltipla.

Eles possibilitam utilizar outras distribuições para os erros e uma função de ligação relacionando a média da variável resposta à combinação linear das variáveis explicativas.

Os modelos lineares generalizados são ajustados aos dados pelo método de máxima verossimilhança, fornecendo não apenas estimativas dos coeficientes de regressão, mas também erros padrão assintóticos estimados, ou seja, em amostras grandes dos coeficientes.

Estima o beta, nas componentes sistemáticas

Modelo satisfatório? o deviance e se os betas são significativos

Na distribuição gama é importante a variabilidade, ou seja o deviance é escalonado

é a partir da componente aleatória que se define qual modelo deve ser utilizado

Funções de ligação: IDENTIDADE, LOGARÍTMICA E RECIPROCIDADE

Regressão linear os dados tem que simétricos

estimativa é com máxima verossimilhança

REGRESSÃO GAMMA

São para dados positivos e assimétricos

o deviance tem que ser o **escalonado** = deviance residual X Alpha do gamma.shape

com isso constrói a tabela pro resíduo o valor tem que cair dentro da área de aceitação, p value alto, acima de 5%

**Regressão Logística**

* É um modelo que tem a análise de qualidade se o modelo satisfatório é curva ROC, Matriz de confusão;
* Distingue-se pelo fato de a variável resposta ser binária Yi=0 (fracasso) e Yi=1(sucesso);
* Método da máxima verossimilhança para estimar os parâmetros (β0 e β1);
* A interpretação dos parâmetros de um modelo de regressão logística é obtida comparando a probabilidade de sucesso com a probabilidade de fracasso, usando a função odds ratio - OR (razão de chances);
* Os dados precisam de distribuição binomial para ajustar um modelo logístico;
* Estimar os parâmetros do modelo: método iterativo de Newton-Raphson.

*Seleção de variáveis:*

A abordagem tradicional na construção de modelos estatísticos é encontrar o modelo mais parcimonioso que explica os dados. Quanto mais variáveis no modelo, maior se torna a estimativa do erro e mais dependente o modelo fica dos dados observados.

Stepwise: procedimento para seleção ou exclusão de variáveis de um modelo é baseado em um algoritmo, AIC e BIC

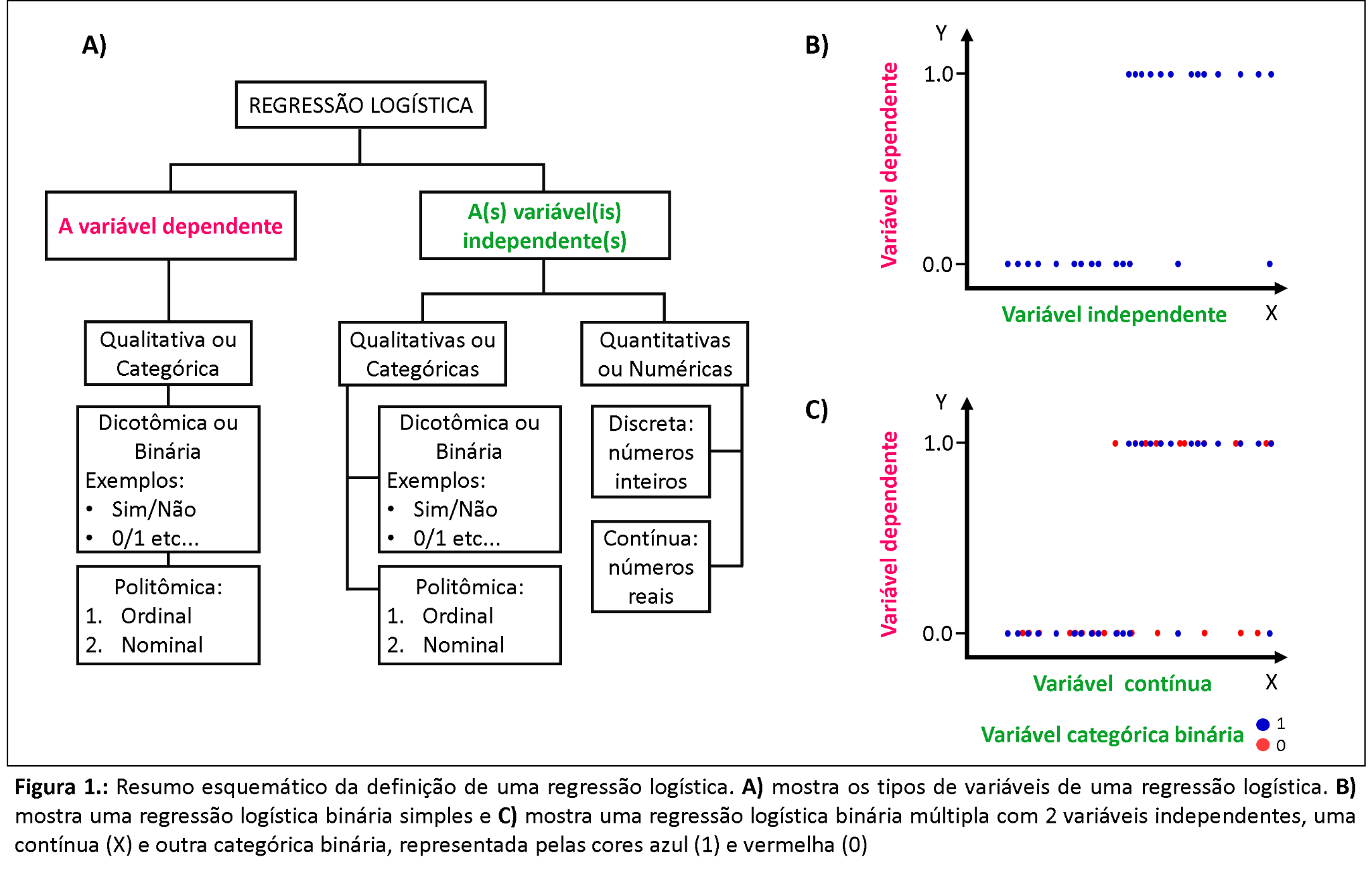
*Ajuste do modelo:*

Teste de Qualidade do Ajuste de Pearson: avalia a discrepância entre o modelo atual e o modelo completo. Determina se as probabilidades preditas se desviam das probabilidades observadas. Se o p-valor do teste for menor do que seu nível de significância escolhido, as probabilidades preditas se desviam das probabilidades observadas.

*Predição:*

É válido nesse momento pensar em definir qual ponto de probabilidade define se o “futuro” ponto pertence à categoria “0” ou “1”. Esse ponto é conhecido como ponto de corte.

ROC: Uma forma bastante utilizada para determinar o ponto de corte é através da Curva ROC (Receiver Operating Characteristic Curve). A curva ROC plota sensibilidade versus 1-especificidade

****

**Binária**

0 ou 1

**Multinomial**

* Quando a variável resposta qualitativa tem mais do que duas categorias, por exemplo, a pressão sanguínea pode ser classificada em: alta, normal e baixa.

**Regressão para Dados de Contagem - Poisson**

**Exemplos:**

* Número de acidentes em uma rodovia por semana;
* Número de automóveis vendidos por dia;

Modelos probabilísticos para variáveis aleatórias quantitativas discretas, com suporte no conjunto de números inteiros não-negativos, são potenciais candidatos para a análise de dados de contagens.

# **Regressão para Dados de Contagem - Binomial Negativa**

**Regressão para Dados Positivos e Assimétricos**

A classe de modelos para a análise de dados positivos assimétricos é bastante ampla, incluindo distribuições conhecidas para os erros, tais como gama, normal inversa, Weibull, Pareto e log-normal.

**Deviance**

Deviance: avalia a discrepância entre o modelo atual e o modelo simples por meio da soma dos quadrados dos resíduos. A estatística D, sob a suposição que o modelo ajustado é correto, tem distribuição assintótica qui-quadrado com (J-p) graus de liberdade. Se o p-valor do teste for menor do que seu nível de significância escolhido, as probabilidades preditas se desviam das probabilidades observadas.

Motivos comuns para o desvio:

- função de ligação incorreta;

- dispersão alta dos dados;

- covariáveis não bem relacionadas com a resposta.

o *residual deviance* tem que ser menor que *Null deviance* Assim, um valor mais baixo do desvio residual indica que o modelo ficou melhor quando inclui as duas variáveis independentes. Já o AIC também é baseado no desvio (*deviance*) e é útil quando queremos comparar modelos e seu valor por si só não é interpretável.

**Odds ration**

https://rpubs.com/StevenDuttRoss/razao\_de\_chances#:~:text=Se%20a%20probabilidade%20de%20sucesso,probabilidade%20aumenta%20ou%20vice%2Dversa.