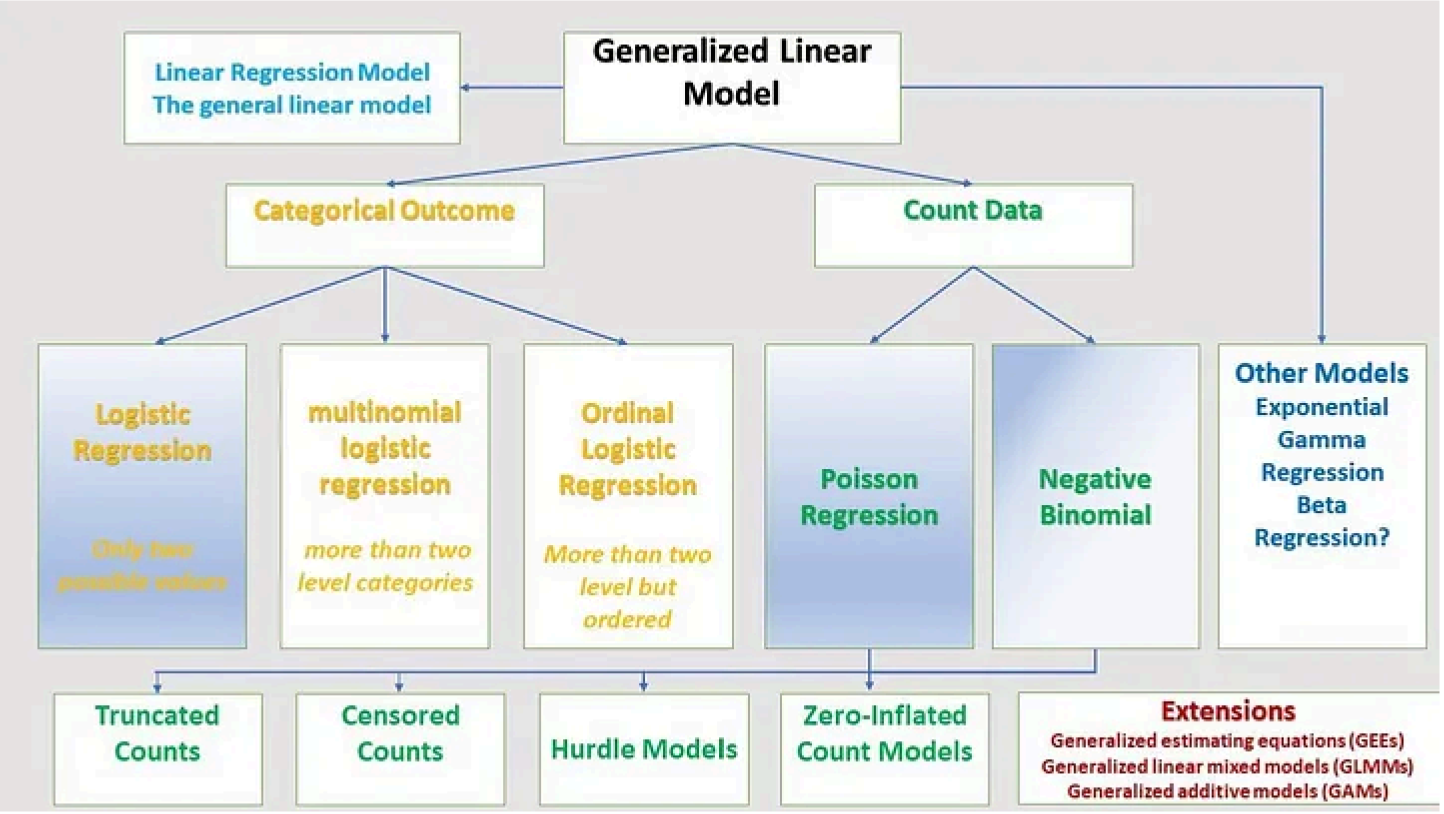
Generalized Linear Models



- This is a diagram **GLMs** Generalized Linear Models which we can notice that all the statistical learning methods we covered so far are are derived from **Linear Regression** by extending it to accommodate various types of response Y variables .

Recall that Linear Regression model rely on these two assumptions :

- Relationship between X and Y is **Linear**
- Normality of Residuals $Y - \hat{Y}$
- Independence of Error
- constant variance of error terms

GLMs try to expand the Linear Regression and make it more flexible by :

- Allow for the response to follow various distributions $\rightarrow$ **Binomial** , **Poisson** , **gamma**
- Allow non-linear relationships within the linear framework
- More flexible on the assumptions and limitations
- Allow for different error structures accommodating hetroscedasticity

Scope of Application

- Linear Regression Models are best suited for **Continuous data** that fits Gaussian distribution
- **GLMs** can handle border scope of data types $\rightarrow$ binary, count and continuous data

Math Behind GLM:

GLMs are based on the assumption of the response Y variable distribution comes from the **exponential family** that includes : Gaussian, Binomial, Poisson and gamma

$$f(y|\theta) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

- y is the response outcome variable
- θ is the parameter as β the coefficient in linear regression
- ϕ Dispersion parameter which is the scales variance
- $a(\phi), b(\theta), c(y,\phi)$ Known functions that define the distribution

Exponential Family of Distributions

Distribution	Domain	Parameter θ	$b(\theta)$	$Var(Y)$	Used For
Gaussian	$(-\infty, \infty)$	μ	$\frac{\theta^2}{2}$	σ^2	Linear Regresion
Binomial	$(0, 1, \ldots, n)$	$\log\left(\frac{p}{1-1}\right)$ Logit	$\log(1+e^\theta)$	$np(1-p)$	Logistic Regresison
Multinomial		$\log\left(\frac{e^{X_i}}{\sum e^{X_j}}\right)$	$\log\left(1+\sum_{k=1}^{K-1} e_k^\theta\right)$		Multi-class Logistic Regression
Poisson	$(0, 1, 2\ldots)$	$\log(\lambda)$	e^θ	λ	Poisson Regression
Exponential	$(0, \infty)$	$-\lambda$	$-\log(-\theta)$	$\frac{1}{\lambda^2}$	Survival Analysis
Gamma	$(0, \infty)$	$-\frac{\nu}{\mu}$	$-\log(-\theta)$	$-\frac{\mu^2}{\nu}$	Positive Continuous Data

GLM's Components

These are the core components that define the Generalized Linear Models structure and functionality

1. Error distribution
2. Linear predictor
3. Link function

Error Distribution

This referred to the probability distribution of the response variables which determines the form of the likelihood function used for the parameters estimations, the selection of the distribution is guided by the nature of the response variables (**continuous, binary, count,...**) The **GLM's** include :

- **Gaussian Distribution** Used for continuous response Y and where the residuals are assumed to be normally distributed
- **Binomial Distribution** Fits the Binary response variables as in Logistic Regression where the outcome is in the form **Two classes**
- **Poisson Distribution** Fits when the response variables represent an event that occur in a fixed interval or space, as in the **Poisson Regression example** discuss later on
- **Gamma Distribution** Used for continuous positive data, often used when the response Y variables represent time until event occurs

Linear Predictor

The linear predictor as heavily discuss in the Linear Regression chapter where it a **Linear combination** of these variables each multiplied by a Coefficient that quantifies the relationship between the predictor and the response variables Y :

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Link Function

The link function connects the linear predictor to the mean μ , It transforms the expected value of the response Y variable to the scale of the linear predictor, to make sure it stay within the range for example **Logistic Regression** logit function that transforms the probability to an unbounded

Distribution	Function	Purpose	Formula
Gaussian	Identity	Directly relates the linear predictor to the response	$\eta = \mu$
Binomial	Logit	Transforms the probability of the success to an unbounded scale	$\eta = \log\left(\frac{\mu}{1-\mu}\right)$
Multinomial	Softmax	Generalizes the logit function for multi class outcome	$\eta_i = \log\left(\frac{e^{X_i}}{\sum e_i^X}\right)$
Poisson	Log	Relates the log of the meant count to the linear predictors, Used for count data	$\eta = \log(\mu)$

- μ is he response variable's distribution

Poisson Regression

This section covers the **Poisson Regression** and its use case when the **Linear Regression** doesn't fit the problem at hand