MAT 3775
Analyse de la régression

Chapitre 7
Modèles de régression généralisée

P. Boily (uOttawa)

Session d'hiver – 2023

Aperçu

7 – Modèles de régression généralisée

Dans les situations réelles, l'hypothèse de linéarité (c'est-à-dire l'hypothèse selon laquelle la réponse ajustée prend la forme $\hat{\mathbf{Y}} = \mathbf{Xb}$) n'est pas toujours respectée (c'est certainement le cas si la réponse doit être non négative ou catégorielle, par exemple).

Il existe diverses approches de modélisation non linéaire pour gérer de telles situations, notamment les réseaux neuronaux, les machines à vecteurs de support, les arbres de décision, etc.

La machinerie des moindres carrés est utile et facile à utiliser ; cela serait un gaspillage de la laisser aller à la dérive. Les modèles linéaires généralisés offrent une flexibilité de modélisation supplémentaire sans nécessairement sacrifier la commodité de la régression linéaire.

7.1 – Cadre général

Considérons un ensemble de données $\{(X_i, Y_i) \mid i = 1, \ldots, n\}$ pour lequel la réponse Y est **catégorielle** à $m = 2$ niveaux ($Y \in \{0, 1\}$?). On écrit

$$P(Y_i = 1) = \beta_0 + \beta_1 X_i = \pi_i,\ \mathrm{E}\{Y_i\} = 1 \cdot \pi_i + 0 \cdot (1 - \pi_1) = \pi_i$$

$$\sigma^2\{Y_i\} = \mathrm{E}\{Y_i^2\} - (\mathrm{E}\{Y_i\})^2 = 1^2 \cdot \pi_i + 0^2 \cdot (1 - \pi_i) - \pi_i^2 = \pi_i(1 - \pi_i)$$

Pour un modèle de RLS $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ avec des erreurs **indépendantes** :

1. l'image de ε ne contient que $m = 2$ valeurs pour un niveau de prédicteur X fixe,

$$\varepsilon = \begin{cases} 1 - (\beta_0 + \beta_1 X), & \text{if } Y = 1 \\ 0 - (\beta_0 + \beta_1 X), & \text{if } Y = 0 \end{cases}$$

2. la variance des termes d'erreur est de

$$\sigma^2\{\varepsilon_i\} = \sigma^2\{Y_i\} = \pi_i(1-\pi_i) = (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i)$$

3. les $E\{Y_i\} = \pi_i$ sont telles que $\pi_i \in [0,1]$.

La première conséquence viole l'hypothèse que les termes d'erreur sont **normaux** ; la deuxième que la variance est **constante**, et la troisième que la réponse est **linéaire** en fonction du prédicteur.

Les deux premiers problèmes peuvent être atténués par le TLC et la régression pondérée si n est suffisamment grand, mais pas le troisième : la RLS n'est **pas** un modèle approprié dans ce cas.

Les **modèles linéaires généralisés** (MLG) étendent le paradigme des moindres carrés ordinaires (OLS) en prenant en compte des réponses qui suivent des lois **non-normales**.

Mis à part la **structure des erreurs**, un MLG est essentiellement un modèle linéaire

$$Y_i \sim \mathcal{D}(\mu_i), \quad \text{où } g(\mu_i) = \eta_i = \mathbf{X}_i\boldsymbol{\beta}.$$

Un MLG est constitué d'une :

- **composante systématique** $\eta_i = \mathbf{X}_i\boldsymbol{\beta}$ (linéaire en $\boldsymbol{\beta}$) ;
- **composante aléatoire**, spécifiée par la distribution $\mathcal{D}$, et
- **fonction de liaison** g.

Les idées et concepts généraux des OLS se reportent sur les MLG :

- la **composante systématique** est spécifiée à l'aide d'un prédicteur linéaire $\eta_i = \mathbf{X}_i\boldsymbol{\beta}$;
- la **fonction de liaison** g associant la **composante systématique** η_i à la **distribution** de la réponse Y_i doit être **lisse** (ou au moins différentiable) et **monotone** (et donc **inversible**) ;
- la **distribution** $\mathcal{D}$ pour la réponse Y_i est généralement choisie dans la **famille exponentielle** des distributions, avec une f.p.d. satisfaisant à

$$f(\mathbf{x} \mid \boldsymbol{\theta}) = h(\mathbf{x})k(\boldsymbol{\theta}) \exp(\boldsymbol{\phi}(\boldsymbol{\theta}) \cdot \boldsymbol{T}(\mathbf{x})), \quad \text{pour de bons } h,\ k,\ \phi,\ \mathbf{T};$$

ceci inclut les lois normale, binomiale, de Poisson, Gamma, etc.

Le modèle des moindres carrés est un exemple d'un MLG, avec :

- **composante systématique** $\eta_i = \mathbf{X}_i\boldsymbol{\beta}$;
- **composante aléatoire** $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$;
- **fonction de liaison** $g(\mu) = \mu$,

que nous avons précédemment exprimé sous la forme

$$\mathbf{Y} \mid \mathbf{X} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n).$$

L'exemple du début de cette section est également un MLG.

Les principaux **avantages** des MLG sont que :

- il n'est pas nécessaire de transformer $\mathbf{Y} \mid \mathbf{X}$ si elle n'est pas normale ;
- si le lien produit **des effets additifs**, l'homoscédasticité n'est pas requise ;
- le choix du lien est distinct du choix de la composante aléatoire ;
- les modèles sont ajustés à l'aide du **maximum de vraisemblance** ;
- les **outils d'inférence** et **vérification du modèle** s'appliquent toujours ;
- ils sont facilement implémentés dans `R` *via* `glm()`, et
- diverses approches de modélisation de la régression (OLS, logistique, etc.) sont des MLG.

7.2 – Régression logistique

Considérons n réponses de Bernoulli $Y_i \in \{0, 1\}$; la **composante aléatoire** du MLG est

$$Y_i \sim \mathcal{B}(\pi_i), \quad \text{(où } \pi_i \text{ joue le rôle de } \mu_i\text{)},$$

avec

$$\mathrm{E}\{Y_i\} = \pi_i = P(Y_i = 1 \mid \mathbf{X}_i) \quad \text{et} \quad \sigma^2\{Y_i\} = \pi_i(1 - \pi_i).$$

Nous définissons la **fonction de liaison** g par

$$g(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right).$$

Avec la **composante systématique** $\eta_i = \mathbf{X}_i\boldsymbol{\beta}$, le modèle de **régression logistique** est

$$g(\pi_i) = \eta_i, \quad Y_i \sim \mathcal{B}(\pi_i), \quad \text{erreurs indép.,}$$

que l'on peut ré-écrire selon

$$\begin{aligned}
g(\pi_i) = \eta_i &\iff \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{X}_i\boldsymbol{\beta} \iff \frac{\pi_i}{1-\pi_i} = \exp(\mathbf{X}_i\boldsymbol{\beta}) \quad (\text{“odds ratio”}) \\
&\iff \pi_i = (1-\pi_i)\exp(\mathbf{X}_i\boldsymbol{\beta}) \\
&\iff \pi_i + \pi_i\exp(\mathbf{X}_i\boldsymbol{\beta}) = \exp(\mathbf{X}_i\boldsymbol{\beta}) \\
&\iff \pi_i = \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}, \quad Y_i \sim \mathcal{B}(\pi_i).
\end{aligned}$$

7.2.1 – Estimation du maximum de vraisemblance

La question qui se pose alors est la suivante : comment trouver le vecteur de paramètres $\boldsymbol{\beta}$? La f.d.p. d'une v.a. de Bernoulli $Y_i \sim \mathcal{B}(\pi_i)$ est

$$f(Y_i) = \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}, \quad Y_i \in \{0,1\};$$

si nous supposons l'indépendance des termes d'erreur, la fonction conjointe de **maximum de vraisemblance** devient

$$L(\boldsymbol{\beta}) = f(\mathbf{Y}) = \prod_{i=1}^{n} f(Y_i) = \prod_{i=1}^{n} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}, \quad Y_i \in \{0,1\}.$$

La **log-vraisemblance logarithmique** est ainsi

$$\begin{aligned}
\ln L(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \ln f(Y_i) = \sum_{i=1}^{n} \left[Y_i \ln \pi_i + (1 - Y_i) \ln(1 - \pi_i)\right] \\
&= \sum_{i=1}^{n} \left[Y_i \ln \pi_i - Y_i \ln(1 - \pi_i) + \ln(1 - \pi_i)\right] \\
&= \sum_{i=1}^{n} Y_i \ln \left(\frac{\pi_i}{1 - \pi_i}\right) + \sum_{i=1}^{n} \ln(1 - \pi_i) \\
&= \sum_{i=1}^{n} Y_i \mathbf{X}_i \boldsymbol{\beta} + \sum_{i=1}^{n} \ln \left(1 - \frac{\exp(\mathbf{X}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta})}\right),
\end{aligned}$$

ou encore

$$\begin{aligned}\ln L(\boldsymbol{\beta}) &= \sum_{i=1}^{n} Y_i \mathbf{X}_i \boldsymbol{\beta} + \sum_{i=1}^{n} \ln\left(\frac{1}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right) \\ &= \sum_{i=1}^{n} Y_i \mathbf{X}_i \boldsymbol{\beta} - \sum_{i=1}^{n} \ln\left(1+\exp(\mathbf{X}_i\boldsymbol{\beta})\right)\end{aligned}$$

Le vecteur de l'estimateur du **maximum de vraisemblance** b satisfait à

$$0 = \left.\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\mathbf{b}} = \nabla_{\boldsymbol{\beta}} \ln L(\boldsymbol{\beta})\big|_{\boldsymbol{\beta}=\mathbf{b}}$$

qui, contrairement au cas des OLS, doit être trouvé en utilisant des **méthodes numériques**.

Les **valeurs ajustées** correspondantes sont

$$\hat{\pi}_i = \frac{\exp(\mathbf{X}_i \boldsymbol{b})}{1+\exp(\mathbf{X}_i \boldsymbol{b})},$$

la **fonction de réponse logistique ajustée** à un $\mathbf{x}$ générique est

$$\hat{\pi}(\mathbf{x}) = \frac{\exp(\mathbf{x}\boldsymbol{b})}{1+\exp(\mathbf{x}\boldsymbol{b})},$$

et la **fonction de réponse logit ajustée** à $\mathbf{x}$ est

$$\hat{\pi}'(\mathbf{x}) = \mathbf{x}\boldsymbol{b} = \ln\left(\frac{\hat{\pi}(\mathbf{x})}{1-\pi(\mathbf{x})}\right).$$

Exemple : nous utilisons un modèle de régression logistique afin d'ajuster le statut de maladie $(Y = 0, 1)$ d'un individu dans une population en fonction de son âge (X_1), de sa situation professionnelle $(X_2 = 0, 1)$, et de sa résidence $(X_3 = 0, 1)$. Les coefficients obtenus sont

$$\boldsymbol{b} = (b_0, b_1, b_2, b_3) = (-4.9988, 0.13086, -0.41636, 2.6858).$$

Quelle est la probabilité que l'individu i ait la maladie $(Y = 1)$ si son vecteur de prédiction est $\mathbf{X}_i = (44, 1, 1)$?

Solution : la fonction de réponse logistique ajustée est

$$\hat{\pi}(\mathbf{x}) = \frac{\exp(-4.9988 + 0.13086x_1 - 0.41636x_2 + 2.6858x_3)}{1 + \exp(-4.9988 + 0.13086x_1 - 0.41636x_2 + 2.6858x_3)},$$

d'où $P(Y = 1 \mid \mathbf{X} = (44, 1, 1)) = \hat{\pi}(44, 1, 1) = 0.9538443$.

Interprétation du coefficient b_k

Soient $\mathbf{X}_1^k, \mathbf{X}_2^k = \mathbf{X}_1^k + \mathbf{e}_k$ dans la portée du modèle, c'est-à-dire que $\mathbf{X}_2^k$ diffère de $\mathbf{X}_1^k$ d'une seule unité à la $k$ième position, i.e., $\Delta X_k = 1$.

D'une part,

$$\Delta\hat{\pi}' = \hat{\pi}'(\mathbf{X}_2^k) - \hat{\pi}'(\mathbf{X}_1^k) = \hat{\pi}'(\mathbf{X}_1^k) + b_k - \hat{\pi}'(\mathbf{X}_1^k) = b_k;$$

D'autre part,

$$\begin{aligned}\Delta\hat{\pi}' = \hat{\pi}'(\mathbf{X}_2^k) - \hat{\pi}'(\mathbf{X}_1^k) &= \ln\left(\frac{\hat{\pi}(\mathbf{X}_2^k)}{1-\hat{\pi}(\mathbf{X}_2^k)}\right) - \ln\left(\frac{\hat{\pi}(\mathbf{X}_1^k)}{1-\hat{\pi}(\mathbf{X}_1^\mathbf{k})}\right) \\ &= \ln\left(\frac{\hat{\pi}(\mathbf{X}_2^k)}{1-\hat{\pi}(\mathbf{X}_2^k)}\Big/\frac{\hat{\pi}(\mathbf{X}_1^k)}{1-\hat{\pi}(\mathbf{X}_1^k)}\right).\end{aligned}$$

Ainsi,

$$e^{b_k} = \frac{\text{odds ratio}(\mathbf{X_2^k})}{\text{odds ratio}(\mathbf{X_1^k})}$$
$$= \frac{\hat{\pi}(\mathbf{X}_2^k)}{1-\hat{\pi}(\mathbf{X}_2^k)} \Bigg/ \frac{\hat{\pi}(\mathbf{X}_1^k)}{1-\hat{\pi}(\mathbf{X}_1^k)};$$

l'augmentation des % des odds ratios avant et après le changement est alors de $e^{b_k} - 1$.

Exemple : supposons que $e^{b_k} = 1.26$; alors si X_k augmente de 1, il y a une augmentation de 26% de l'odds ratio.

Si $e^{b_k} = 0.25$, alors si X_k augmente de 1, il y a au contraire une chute de 75% de l'odds ratio.

7.2.2 – Signification des prédicteurs

Comme dans le cas des OLS, nous pouvons fournir des tests inférentiels pour les paramètres du modèle : lorsque n est élevé, on utilise la statistique de test

$$z = \frac{b_k - \beta_k}{s\{b_k\}}.$$

Mais qu'est-ce que $s\{b_k\}$ dans le contexte de la **régression logistique** ?

Nous utilisons le **test de Wald** pour tester

$$\begin{cases} H_0 : \beta_k = 0 \\ H_1 : \beta_k \neq 0 \end{cases}$$

Si H_0 est valide, la statistique de test est $z^* = \frac{b_k}{s\{b_k\}}$.

L'écart-type est fourni par la **matrice approximative de variance-covariance estimée**

$$\mathrm{s}^2\{\mathbf{b}\} = -G^{-1} = -\left[\frac{\partial^2}{\partial\beta_i\partial\beta_j}\ln L(\beta)\right]^{-1},$$

qui doit une fois de plus être calculée numériquement. Lorsque n est élevé et que H_0 est valide, on s'attend à ce que $z^* \sim \mathcal{N}(0,1)$, approximativement.

Règle de décision : si $|z^*| > z(1-\frac{\alpha}{2})$, on **rejette** H_0 à un niveau de confiance α ; $b_k \pm z(1-\frac{\alpha}{2})s\{b_k\}$ est un I.C. de β_k à environ $100(1-\alpha)\%$.

On peut aussi effectuer des inférences simultanées (cf. le module 3).

Exemple : supposons que la matrice de variance-covariance approximative de l'exemple de l'état de la maladie est la suivante

$$\mathrm{s}^2\{\mathbf{b}\} = \begin{pmatrix} 0.5130 & -0.1168 & -0.3121 & -0.2743 \\ -0.1168 & 0.00029 & 0.00084 & 0.00045 \\ -0.3121 & 0.00084 & 0.4761 & 0.1734 \\ -0.2743 & 0.00045 & 0.1734 & 0.3627 \end{pmatrix}.$$

Trouver un $\mathrm{I.C.}$ de β_3 et pour l'odds ratio $\exp(\beta_3)$, à environ 95%.

Solution : nous avons $\alpha = 0.05$, $b_3 = 2.6858$, $s\{b_3\} = \sqrt{0.3627} = 0.6022$, et $z(1-0.05/2) = z(0.975) = 1.96$. Ainsi

$$\text{CI}(\beta_3; 0.95) : 2.6858 \pm 1.96(0.6022) \equiv (1.505, 3.866) \quad \text{et}$$

$$\text{CI}(\exp(\beta_3); 0.95) : \exp(1.505, 3.866) \equiv (4.506, 47.756).$$

7.2.3 – Test d'adéquation de la déviance

Il n'existe **aucune** statistique descriptive facilement disponible qui agisse pour la régression logistique comme R^2 ou R_a^2 le font pour les OLS (**des candidats existent, tels que la R^2 de Nagelkerke, mais ils ont tous des défauts importants**).

Nous utilisons au lieu le **test d'adéquation de la variance** :

$$\begin{cases} H_0 : \pi = \dfrac{\exp(\mathbf{X}\beta)}{1+\exp(\mathbf{X}\beta)} \\ H_1 : H_0 \text{ n'est pas valide} \end{cases}$$

On peut l'utiliser si l'on dispose de mesures répétées pour **au moins un** des niveaux de prédiction (autres tests : χ^2 de Pearson, Hosmer-Lemeshow).

Soient c différents niveaux de prédicteurs $(\mathbf{X}_1, \ldots, \mathbf{X}_c)$ et la $i$ème réponse du groupe j $Y_{i,j}$ (i.e., lorsque $\mathbf{X} = \mathbf{X}_j$), $i = 1, \ldots, n_j$, $j = 1, \ldots, c$.

Les modèles **complet** et **réduit** sont $\mathrm{E}\{Y_{i,j}\} = \pi_j$, $\mathrm{E}\{Y_{i,j}\} = \hat{\pi}(\mathbf{X}_j) = \hat{\pi}_j$. Soit p_j la **proportion des observations avec $Y_{i,j} = 1$ dans le groupe j**.

La statistique de la **déviance du modèle ajusté** est

$$\mathrm{DEV}(p) = -2\sum_{j=1}^{c}\left[n_j p_j \ln\frac{\hat{\pi}_j}{p_j} + (n_j - n_j p_j)\ln\left(\frac{1-\hat{\pi}_j}{1-p_j}\right)\right],$$

qui suit une loi $\chi^2(c-p)$ lorsque H_0 est valide.

Règle de décision : si $\mathrm{DEV}(p) > \chi^2(1-\alpha; c-p)$, on **rejette** H_0 à un niveau de confiance α.

Exemple : nous ajustons un modèle logistique avec $p = 3$ afin d'expliquer si $c = 6$ groupes d'étudiants réussiront un examen difficile du premier coup. Les statistiques d'ajustement sont présentées ci-dessous :

groupe j	n_j	$\hat{\pi}_j$	p_j	**groupe** j	n_j	$\hat{\pi}_j$	p_j
1	190	0.1865	0.154	4	202	0.4920	0.515
2	186	0.2645	0.296	5	175	0.7718	0.742
3	200	0.3562	0.350	6	112	0.8557	0.902

Le modèle logistique correspondant est-il acceptable, lorsque $\alpha = 0.05$?

Solution : Nous avons $p = 3$ et

$$\mathrm{DEV}(3) = -2\left[n_1p_1 \ln\left(\frac{\hat{\pi}_1}{p_1}\right) + (n_1 - n_1p_1)\ln\left(\frac{1-\hat{\pi}_1}{1-p_1}\right) + \cdots\right.$$

$$\left.\cdots + n_6p_6 \ln\left(\frac{\hat{\pi}_6}{p_6}\right) + (n_6 - n_6p_6)\ln\left(\frac{1-\hat{\pi}_6}{1-p_6}\right)\right]$$

$$= -2\left[190(0.154)\ln\left(\frac{0.1865}{0.154}\right) + (190 - 190(0.154))\ln\left(\frac{1-0.1865}{1-0.154}\right) + \cdots\right.$$

$$\left.\cdots + 112(0.902)\ln\left(\frac{0.8557}{0.902}\right) + (112 - 112(0.902))\ln\left(\frac{1-0.8557}{1-0.902}\right)\right] = 5.786.$$

Lorsque $\alpha = 0.05$ et $c - p = 3$, le seuil décisionnel est $\chi^2(0.95; 3) = 7.81$. Comme $\text{DEV}(3) = 5.786 \leq 7.81$, nous concluons que le modèle logistique fournit un ajustement **satisfaisant**.

En pratique, si $p_j = 0, 1$ pour un groupe j, on utilise la convention

$$n_j p_j \ln\left(\frac{\hat{\pi}_j}{p_j}\right) = 0 \quad \text{ou} \quad (n_j - n_j p_j)\ln\left(\frac{1-\hat{\pi}_j}{1-p_j}\right) = 0,$$

respectivement.

7.3 – Régression de Poisson

Lors de la propagation d'une rumeur, la vitesse à laquelle de nouveaux individus apprennent l'information (au début) augmente exponentiellement avec le temps. Si μ_i représente le **# attendu de personnes qui ont l'info au temps** x_i, on utilise un modèle de la forme $\mu_i = \gamma \exp(\delta x_i)$:

$$\underbrace{\ln(\mu_i)}_{\text{liaison}} = \ln \gamma + \delta x_i = \beta_0 + \beta_1 x_i = \underbrace{(1, x_i)^\top (\beta_0, \beta_1)}_{\text{composante systématique}} = \mathbf{x}_i \boldsymbol{\beta}.$$

De plus, puisque nous mesurons un nombre d'individus, la distribution **Poisson** pourrait être un choix raisonnable, conduisant au MLG suivant :

$$Y_i \sim \underbrace{\text{Poisson}(\mu_i),}_{\text{composante aléatoire}} \quad \ln(\mu_i) = \mathbf{x}_i \boldsymbol{\beta}.$$

En général, si $Y_i \sim \text{Poisson}(\mu_i)$ sont indépendants et si la fonction de liaison est $g(\mu_i) = \mathbf{x}_i\boldsymbol{\beta}$, la **fonction de maximum de vraisemblance** est

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(Y_i) = \prod_{i=1}^{n} \frac{e^{-\mu_i}\mu_i^{Y_i}}{Y_i!};$$

la **log-vraisemblance** est

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} Y_i \ln \mu_i - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} \ln(Y_i!),$$

et l'estimateur $\mathbf{b}_{\text{MLE}}$ résoud $\nabla_{\boldsymbol{\beta}} \ln L(\boldsymbol{\beta}) = 0$ (calcul numérique).

Les **valeurs ajustées** correspondantes sont $\hat{\mu}_i = g^{-1}(\mathbf{x}_i\boldsymbol{b})$.

Test du rapport de vraisemblance : soit $q < p$. Nous testons la signification de certains des prédicteurs (possiblement ré-ordonnés)

$$\begin{cases} H_0 : \beta_q = \beta_{q+1} = \cdots = \beta_{p-1} = 0 \\ H_1 : \text{one of } \beta_q, \beta_{q+1}, \ldots, \beta_{p-1} \neq 0 \end{cases}$$

à l'aide de la statistique de test

$$G^2 = -2\ln\left(\frac{L(R)}{L(F)}\right) = -2\ln\left(\frac{L(\mathbf{b}_R)}{L(\mathbf{b}_F)}\right),$$

où $\mathbf{b}_F, \mathbf{b}_R$ sont les estimateurs des modèles **complet** et **réduit**, respectivement, qui suit une loi $\chi^2(p-q)$ lorsque H_0 est valide.

Règle de décision : si $G^2 > \chi^2(1-\alpha; p-q)$, on **rejette** H_0 à un niveau de confiance α.

Test d'adéquation de la déviance: afin de tester pour

$$\begin{cases} H_0 : g(\mu) = \mathbf{X}\beta = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1} \\ H_1 : H_0 \text{ n'est pas valide} \end{cases}$$

on calcule la statistique de test de la **déviance du modèle**

$$G^2 = \mathrm{DEV}(p) = -2 \left(\sum Y_i \ln \frac{\hat{\mu}_i}{Y_i} + \sum (Y_i - \hat{\mu}_i) \right),$$

qui suit une loi $\chi^2(n-p)$ lorsque H_0 est va;ode.

Règle de décision : si $\mathrm{DEV}(p) > \chi^2(1-\alpha; n-p)$, on **rejette** H_0 à un niveau de confiance α.