Ex. seuil, intérêts, particularités

1. $\boxed{\text{fct de lien dérivée}}$ du lien logit

2. $$\left.\begin{array}{r}\text{I.C. autour de la } DE_{50} \\ \text{des proba.}\end{array}\right\} \rightarrow \begin{array}{c}\text{approximation} \\ \text{ou} \\ \boxed{\text{Max. de vrais.}}\end{array}$$

3. $\boxed{\text{Surdispersion}}$

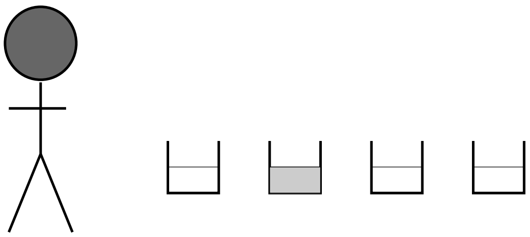
   $$\rightarrow \begin{array}{r}\text{tenir compte des effets séance} \\ \text{et individu}\end{array}$$

4. $\boxed{\text{“efficacité relative”}}$

Exemple seuil

⇒ Déterminer un seuil de détection

Test à choix forcé, 1 parmi 4



Q : Lequel diffère ?

$$\left\langle \begin{array}{lll} Y = 0 & \quad \text{PERDU} & ! \\ Y = 1 & \quad \text{BRAVO} & ! \end{array} \right.$$

Ex. seuil, modèle, fct de lien

$$p = P(Y=1) \qquad \begin{cases} \text{détection réelle} & \Rightarrow \ p_c \text{ (corrigée)} \\ \text{hasard} & \Rightarrow \ \pi_0 = 1/4 \end{cases}$$

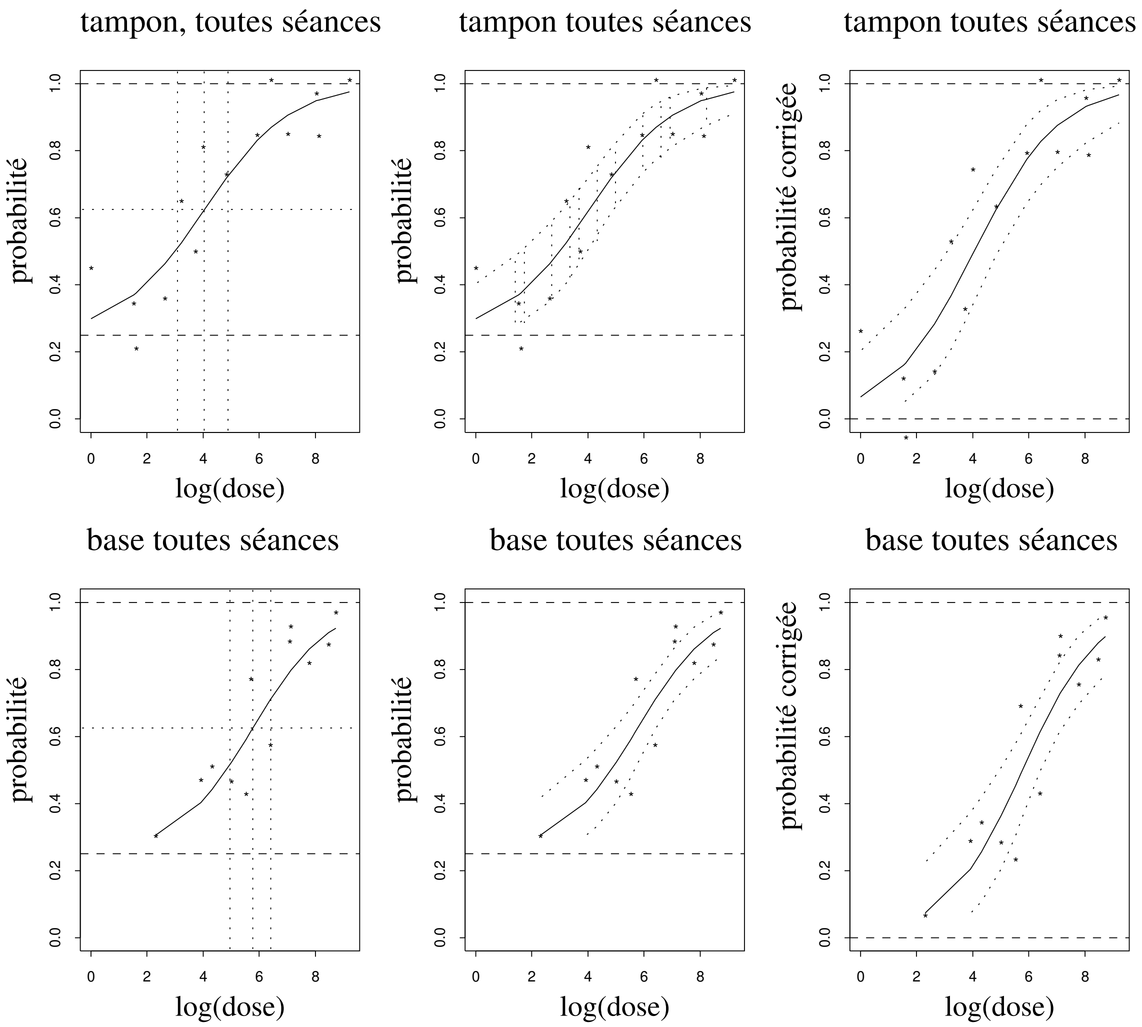
$$\boxed{p = p_c + (1 - p_c)\pi_0}$$

$$\boxed{\text{logit}(p_c) = \log(\frac{p_c}{1-p_c}) = \alpha + \beta \log(conc)}$$

or, $p_c = \frac{p-\pi_0}{1-\pi_0}$, $1 - p_c = \frac{1-p}{1-\pi_0}$

$$\boxed{\log(\frac{p-\pi_0}{1-p}) = \alpha + \beta \log(conc)}$$

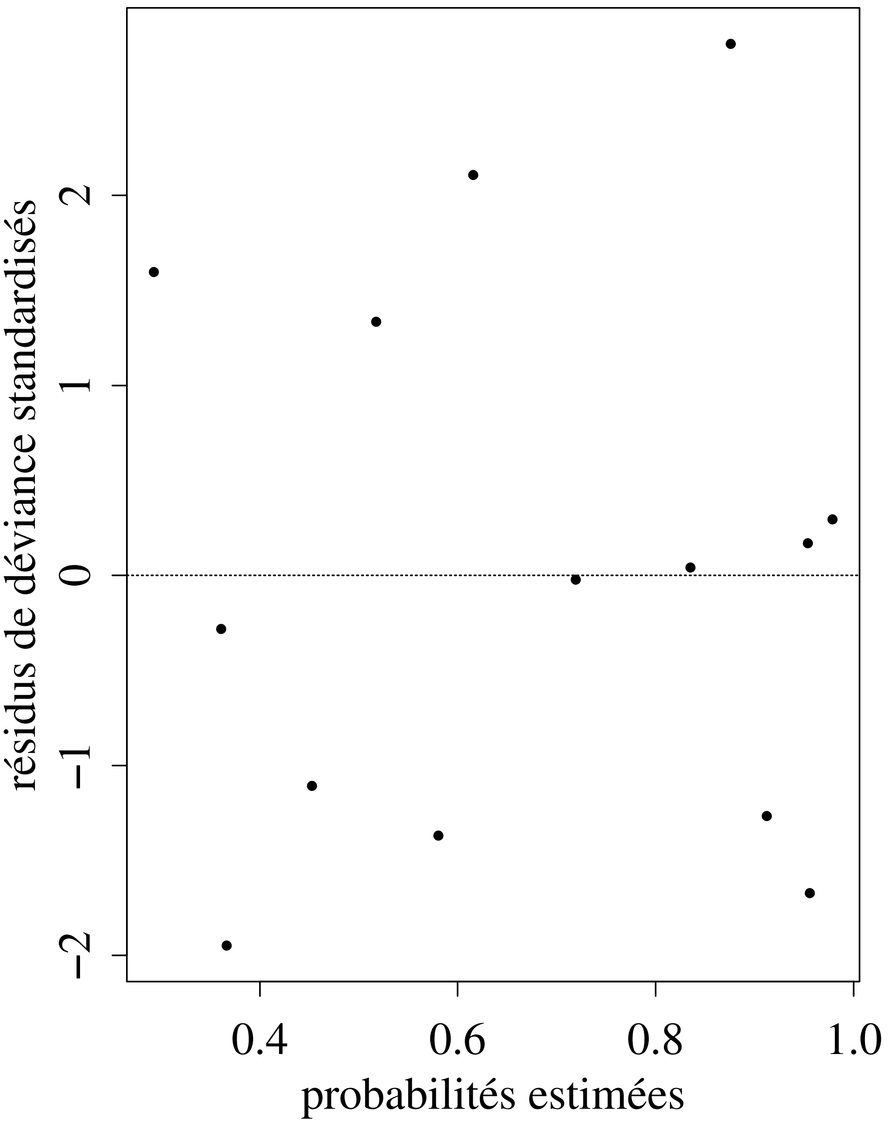
Ex. seuil, résultats



Surdispersion, introduction

```
Deviance Residuals:
  Min   1Q Median 3Q Max
 -1.8 -1.1 0.0085  1 2.6

Coefficients:
                    Std.
            Value  Error tvalue
(Intercept) -2.80   0.71   -3.9
  log(dose)  0.69   0.15    4.7

(Dispersion Parameter
  taken to be 2 )

   Null Deviance: 104 on 13 df

Residual Deviance: 24 on 12 df

Number of
Fisher Scoring Iterations: 4
```

Surdispersion, introduction



Surdispersion, introduction

Loi binomiale

$Y \sim B(n,p)$ $E(Y)=np$ $\underbrace{V(Y)=np(1-p)}_{\text{si les n obs. indépendantes}}$

Si dépendance, variance $\neq n\,p\,(1-p)$

~~binomiale~~

obs. binaires corrélées positivement → variance $> n\,p\,(1-p)$ surdispersion

obs. binaires corrélées négativement → variance $< n\,p\,(1-p)$ sousdispersion

Surdispersion, diagnostic

approx. χ^2 de la déviance pas valide

– si données binaires (non groupées)

– si certains des n_i petits.

---

déviance grande si :

- modèle inapproprié
  - $\rightarrow$ var. expl. oubliée
  - $\rightarrow$ var. expl. à transformer
  - $\rightarrow$ outlier
  - $\rightarrow$ fct de lien inadéquate
- $\boxed{\text{surdispersion}}$

$$\boxed{\text{Surdispersion} \Rightarrow \text{déviance plus grande}}$$

Surdispersion, interprétation intuitive

Surdispersion

2 interprétations possibles

équivalentes

obs. binaires corrélées	un facteur de variation négligé

variation extra-binomiale

Surdispersion, interprétation, hétérogénéité

$\boxed{\text{Variabilité des proba. de réponse}}$

ex. facteur portée

obs : y_i/n_i, $i = 1, \ldots, n$, k var. expl. $X_1 \ldots X_k$

ϑ_i : vraie proba. inconnue de la portée,

varie autour de p_i.

$\hookrightarrow$ V.A.

$$E(\vartheta_i) = p_i$$

$$V(\vartheta_i) = \gamma . p_i . (1 - p_i)$$

γ : param. d'échelle inconnu

Surdispersion, interprétation, hétérogénéité

ϑ_i : V.A non observable : variable latente.

$$y_i|\vartheta_i \sim \mathcal{B}(n_i\vartheta_i, n_i\vartheta_i(1-\vartheta_i))$$

$$E(y_i|\vartheta_i) = n_i\vartheta_i$$

$$V(y_i|\vartheta_i) = n_i\vartheta_i(1-\vartheta_i)$$

Surdispersion, interprétation, hétérogénéité

$$
\begin{aligned}
E(Y) &= E\left[E(Y|X)\right] \\
V(Y) &= E[V(Y|X)] + V[E(Y|X)]
\end{aligned}
$$

$$
\begin{aligned}
E(y_i) &= E[E(y_i|\vartheta_i)] = E(n_i\vartheta_i) = n_i E(\vartheta_i) = n_i p_i \\
V(y_i) &= E[V(y_i|\vartheta_i)] + V[E(y_i|\vartheta_i)]
\end{aligned}
$$

$$
\begin{aligned}
E[V(y_i|\vartheta_i)] &= E[n_i\vartheta_i(1-\vartheta_i)] \\
&= n_i(E(\vartheta_i) - E(\vartheta_i^2))
\end{aligned}
$$

$$\text{or, } V(\vartheta_i) = E\left[(\vartheta_i)^2\right] - \left[E(\vartheta_i)\right]^2$$

$$
\begin{aligned}
E[V(y_i|\vartheta_i)] &= n_i(E(\vartheta_i) - V(\vartheta_i) - \left[E(\vartheta_i)\right]^2) \\
&= n_i(p_i - \gamma\ p_i\ (1-p_i) - p_i^2) \\
&= n_i p_i (1-p_i)(1-\gamma)
\end{aligned}
$$

$$V[E(y_i|\vartheta_i)] = V(n_i\vartheta_i) = n_i^2 V(\vartheta_i) = n_i^2 \gamma\ p_i\ (1-p_i)$$

$$\Rightarrow \boxed{V(y_i) = n_i\ p_i\ (1-p_i)\ \underbrace{(1+(n_i-1)\ \gamma)}_{\phi}}$$

Surdispersion, interprétation, corrélation entre réponses binaires

$y_i/n_i \qquad R_{i1},\ R_{i2},\ \ldots,R_{in_i} \quad : V.A.$binaires $\sim$ Bernouilli

$$P(R_{ij}=1)=p_i$$

$$E(R_{ij})=p_i, \quad V(R_{ij})=p_i\,(1-p_i)$$

y_i : valeur observée de la variable $\sum_j R_{ij}$

$$E(y_i)=\sum_j E(R_{ij})=n_i\,p_i$$

$$V(y_i)=\sum_j V(R_{ij})+\sum_j\sum_{k\neq j} Cov(R_{ij},R_{ik})$$

On pose $\delta = Corr(R_{ij},R_{ik})$

On a $\delta = \dfrac{Cov(R_{ij},R_{ik})}{\sqrt{V(R_{ij}).V(R_{ik})}}$

$$V(R_{ij})=V(R_{ik})=p_i\,(1-p_i)$$

$$Cov(R_{ij},R_{ik})=\delta\,p_i\,(1-p_i)$$

Surdispersion, interprétation, corrélation entre réponses binaires

$$
\begin{aligned}
V(y_i) &= \sum_j p_i\,(1-p_i) + \sum_j \sum_{k \neq j} \delta\, p_i\,(1-p_i) \\
&= n_i\, p_i\,(1-p_i) + n_i\,(n_i-1)\,\delta\, p_i\,(1-p_i)
\end{aligned}
$$

$$
\boxed{V(y_i) = n_i\, p_i\,(1-p_i)\ \boxed{(1+(n_i-1)\,\delta)}}
$$

Surdispersion, interprétation, hétérogénéité et corrélation

$$\boxed{V(y_i) = n_i\, p_i\, (1 - p_i)\, (1 + (n_i - 1)\gamma)} \qquad \mathbf{1}$$

$$\boxed{V(y_i) = n_i\, p_i\, (1 - p_i)\, (1 + (n_i - 1)\,\delta)} \qquad \mathbf{2}$$

1 hétérogénéité : $\gamma \geq 0 \quad \Rightarrow$ surdispersion
2 corrélation : $-1 \leq \delta \leq 1 \Rightarrow \delta \geq 0$ surdispersion
$\delta \leq 0$ sousdispersion

$$(1 + (n_i - 1)\delta) \geq 0 \quad \Rightarrow \quad -\frac{1}{n_i - 1} \leq \delta \leq 1$$

$$\text{limite}: -\tfrac{1}{n_i - 1} \leq \delta$$

Si n_i gd, **2** ne permet pas de modéliser une sousdispersion.

Surdispersion, modélisation, facteur d'hétérogénéité

Modéliser la surdispersion

$$V(y_i) = n_i \, p_i \, (1 - p_i)\phi$$

$$\phi = \sigma^2$$

Si les $n_i \simeq Ct$,

$$\sigma_i^2 = \sigma^2 = \text{facteur d'hétérogénéité}$$

$$\widehat{\sigma}^2 = \frac{\text{déviance (modèle complet)}}{ddl}$$

inférence :

- $e.t.(\widehat{\beta}) \rightarrow$ multiplié par $\sqrt{\widehat{\sigma}^2}$
- $\frac{\widehat{\beta}}{\widehat{e.t.(\widehat{\beta})}} \sim t(ddl)$
- Modèles emboîtés comparés par test F

$$\frac{D_1 - D_0}{(dl_1 - dl_0)\widehat{\phi}} \sim F(dl_1 - dl_0, m - k)$$

Surdispersion, modélisation, méthode de William pour estimer ϕ

principe :

- Modèle complet :

$$\begin{cases} \text{on égalise le } \chi^2 \text{ de Pearson à son espérance,} \\ \text{mais le } \chi^2 \text{ dépend de } \gamma \end{cases}$$

$\hookrightarrow$ méthode itérative $\Rightarrow \widehat{\gamma}$

- Modèles ajustés, avec pondération par $w_i = \frac{1}{\widehat{\sigma_i}^2}$, avec $\widehat{\sigma_i}^2 = 1 + (n_i - 1)\widehat{\gamma}$

Inférence :

- $\frac{\widehat{\beta}}{e.t.(\widehat{\beta})} \sim t(ddl)$
- Modèles emboîtés comparés par test du χ^2

$$(D_1 - D_2) \sim \chi^2(dl_1 - dl_2)$$

Surdispersion, des méthodes pour la prendre en compte

- p_i varie, pas de loi précisée

  $\rightarrow$ facteur de surdispersion ($n_i \simeq \mathrm{Ct}$)

  $\rightarrow$ méthode de William

- $p_i \sim beta$

  $\rightarrow$ modèle beta-binomial

- effets $\sim \mathcal{N}$

  $\rightarrow$ modèles à effets aléatoires

I.C. pour une fonction des paramètres

- ex. 1 : **I.C. pour la DL_{50}**

$$\log(\frac{p-\pi_0}{1-p}) = \alpha + \beta \log(\text{conc})$$

$$\log(\text{conc}) = \boxed{\frac{\log(\frac{p-\pi_0}{1-p}) - \alpha}{\beta}}$$

avec $p = 0.5$ et $\pi_0 = 1/4$

- ex. 2 : **I.C. pour une proba. $\widehat{p_0}$**

I.C. pour $\widehat{\eta_0} = \widehat{\alpha} + \widehat{\beta} \log(\text{conc}_0)$,
puis lien réciproque.

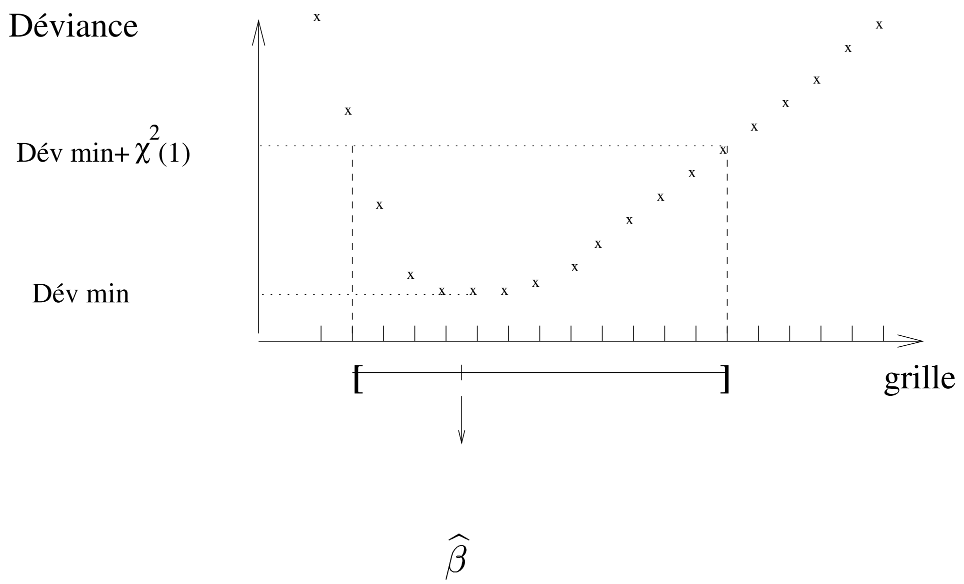
I.C. pour une fonction des paramètres

I.C. pour une fct des param. → approx. normale → I.C. symétrique

I.C. pour une fct des param. → $\boxed{\text{Max. de vrais.}}$ → I.C. dissymétrique

cas particulier :
I.C. pour le paramètre β, par max. de vrais.

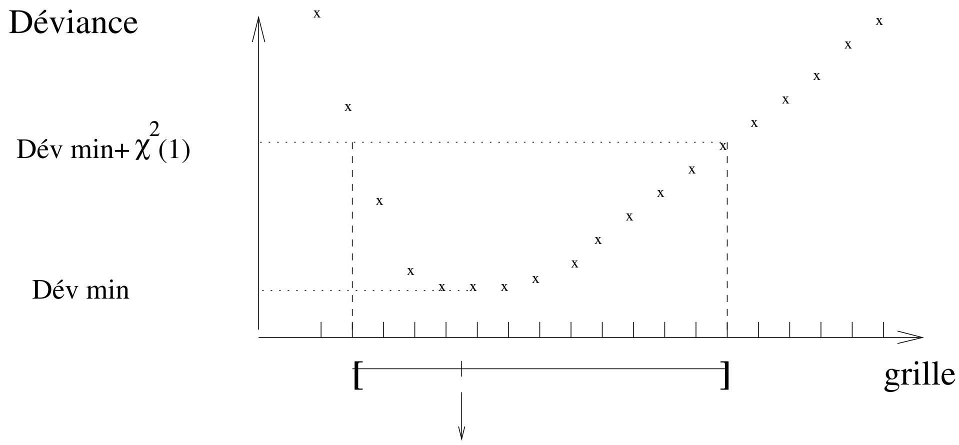
$$\text{logit}(p) = \alpha + \underbrace{\beta}_{grille} \log(conc)$$



I.C. pour une fonction des paramètres

ex. 1 : I. C. pour $\left.\dfrac{\log(\frac{p-\pi_0}{1-p})-\alpha}{\beta}\right|_{p=0.5}$

$$\beta \left[\log(\text{conc}) - \underbrace{\frac{\log(\frac{p-\pi_0}{1-p}) - \alpha}{\beta}}_{grille}\right] + \underbrace{\log(\frac{p-\pi_0}{1-p})}_{offset} = \alpha + \beta \ \log(\text{conc})$$



$$\left.\frac{\log(\frac{p-\pi_0}{1-p}) - \widehat{\alpha}}{\widehat{\beta}}\right|_{p=0.5}$$

I.C. pour une fonction des paramètres

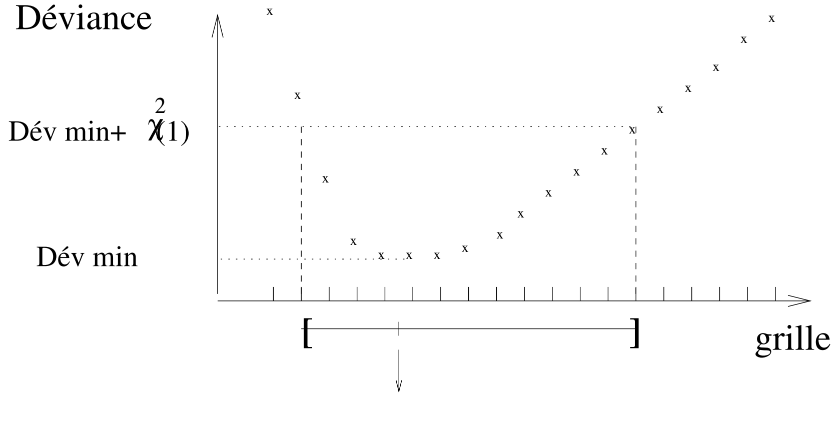
<u>ex. 2 :</u> I. C. pour $\widehat{p_0}$, pour la concentration conc_0

$$\log\left(\frac{p-\pi_0}{1-p}\right) = \alpha + \beta \log \mathrm{conc}$$

I.C. pour $\alpha + \beta \log \mathrm{conc}_0$?

$$\alpha + \beta \log \mathrm{conc}_0 + \beta(\log \mathrm{conc} - \log \mathrm{conc}_0) = \alpha + \beta \log \mathrm{conc}$$

$$\underbrace{\alpha + \beta \log \mathrm{conc}_0}_{offset_{\hookrightarrow grille}} + \beta \ \log \frac{\mathrm{conc}}{\mathrm{conc}_0} = \alpha + \beta \log \mathrm{conc}$$

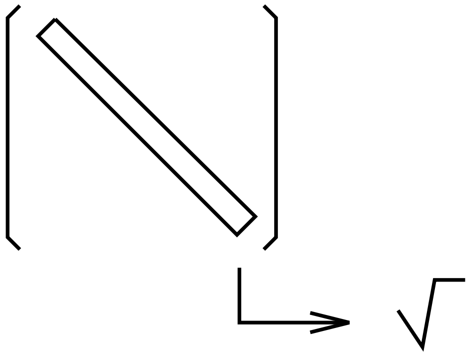


I.C. pour une comb. lin. des params (par approximation)

$$\widehat{\eta}_0 = X_0\widehat{\theta} = \begin{pmatrix} \cdots \\ \cdots \\ \boxed{\cdots\cdots\cdots\cdots} \\ \cdots \end{pmatrix} \begin{pmatrix} \widehat{\theta}_0 \\ \widehat{\theta}_1 \\ \vdots \\ \widehat{\theta}_p \end{pmatrix}$$

$$Var(X_0\,\widehat{\theta}) = X_0\,Var(\widehat{\theta})\;X_0'$$

=



$$[\;\widehat{\eta}_0 \pm z_{\alpha/2}\sqrt{Var(\widehat{\eta}_0)}\;]$$

$$logit^{-1}\;[\;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\;]$$

RÉSUMÉ

Surdispersion (Sousdispersion)
obs. bin. corrélées positivement (négativement)

Si $n_i \simeq Ct$,

$$\begin{aligned} \sigma^2 &= \text{facteur d'hétérogénéité} \\ \widehat{\sigma}^2 &= \frac{\text{Dev(modèle complet)}}{\text{ddl}} \end{aligned}$$

$e.t.(\widehat{\beta}) \longrightarrow$ multipliés par $\sqrt{\widehat{\sigma}^2}$
$\frac{\widehat{\beta}}{\widehat{e.t.(\widehat{\beta})}} \sim t(ddl)$
modèles emboîtés comparés par tests F.

I.C. pour une fct des param. $\nearrow$ approx. normale
I.C. pour une fct des param. $\searrow$ Max. de vrais.