Loi multinomiale

Binomiale : ex. malade Y=O, non malade Y=N

non groupées

	N	O
individus	1	0
	0	1
	⋮	⋮

groupées

	N	O
groupes	10	5
	8	7
	⋮	⋮

Multinomiale : ex. notation maladie

non groupées

	0	1	2	3
individus	0	1	0	0
	1	0	0	0
	⋮	⋮	⋮	⋮

groupées

	0	1	2	3
groupes	9	1	15	17
	5	0	10	13
	⋮	⋮	⋮	⋮

Cas particulier de la loi binomiale

Modèle logit (= Modèle logistique)

$$p_1 = P(Y = 1)$$

$$\log\left(\frac{p_1}{p_0}\right) = \log\left(\frac{p_1}{1 - p_1}\right) = \alpha + \beta x$$

$$\iff p_1 = \frac{exp(\alpha + \beta x)}{1 + exp(\alpha + \beta x)}$$

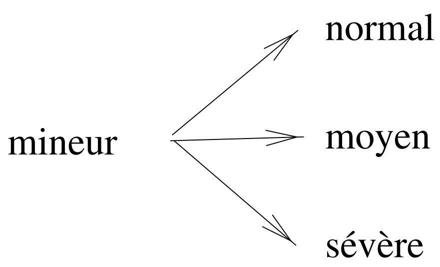
$$\iff p_1 = F(\alpha + \beta x)$$

$$\iff F^{-1}(p_1) = \alpha + \beta x$$

F : fct de répartition de la loi logistique

F^{-1} : fct de lien logit : g.

Exemple mineurs



nb d'années	normal	moyen	sévère
5.8	98	0	0
15.0	51	2	1
21.5	34	6	3
27.5	35	5	8
33.5	32	10	9
39.5	23	7	8
46.0	12	6	10
51.5	4	2	5

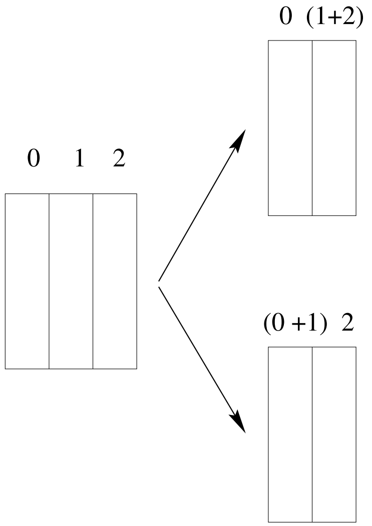
Cas multinomial

$q = 3$: nombre de modalités

$$p_0 = P(Y = 0)$$

$$p_1 = P(Y = 1)$$

$$p_2 = P(Y = 2)$$



$$p_0 + p_1 + p_2 = 1$$

Modèle multinomial

ex. 3 modalités, 1 variable explicative

<u>2 modèles</u> :

$$\log(\frac{p_1 + p_2}{p_0}) \;=\; \alpha_1 + \beta_1\, x$$
$$\log(\frac{p_2}{p_0 + p_1}) \;=\; \alpha_2 + \beta_2\, x$$

<u>1 modèle</u> : $\boxed{\text{MIEUX}}$

$$\begin{cases} \log(\frac{p_1}{p_0}) = \alpha_1 + \beta_1 x \\ \log(\frac{p_2}{p_0}) = \alpha_2 + \beta_2 x \end{cases}$$

$$\Leftrightarrow \log(\frac{p_r}{p_0}) = \alpha_r + \beta_r\, x \qquad\qquad r = 1, 2$$

$$p_0 + p_1 + p_2 = 1$$

$\boxed{\text{Modèle Logit Multinomial}}$

Catégories non ordonnées

Modèle Logit Multinomial, q modalités

$$\log(\frac{p_r}{p_0}) = \alpha_r + \beta_r x, \qquad r = 1, 2$$

$$\Leftrightarrow \; p_r = \frac{exp(\alpha_r + \beta_r x)}{1 + \sum_{s=1}^{q-1} exp(\alpha_s + \beta_s x)}$$

$$\Leftrightarrow \; p_r = F(\alpha_r + \beta_r x)$$

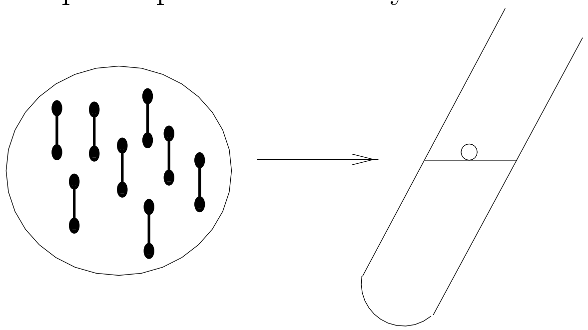
F : fct de répartition de la loi logistique

F^{-1} : fct de lien logit.

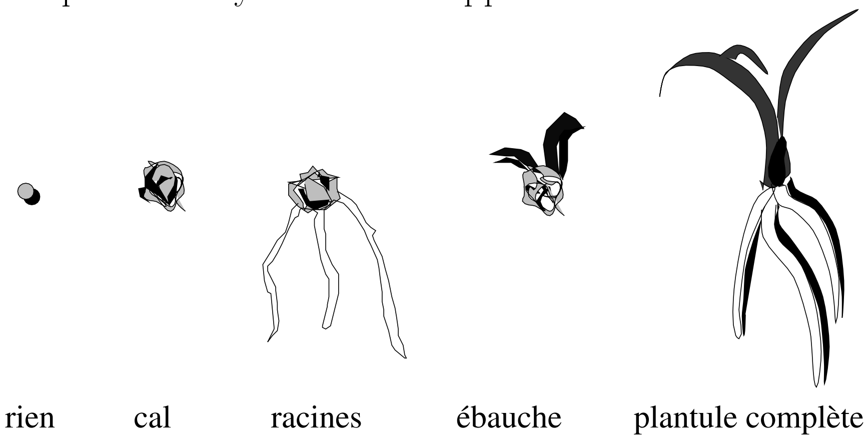
Exemple triticale

Culture in vitro de triticale, protocole

Étape 1 : pollen $\rightarrow$ embryons



Étape 2 : embryons $\rightarrow$ développement



Données brutes

exemple triticale

ligne	Mbase	Kin	Géno	plante	ébauche	racine	cal	rien	n_i
1	1	0	1	20	10	4	41	5	80
2	1	0	2	16	10	9	38	7	80
3	1	0	3	38	7	5	27	3	80
4	1	0	4	10	7	14	22	17	70
5	1	1	1	16	9	15	33	7	80
6	1	1	2	3	11	3	26	37	80
7	1	1	3	13	7	5	35	20	80
8	1	1	4	3	7	6	33	21	70
9	2	0	1	21	8	9	39	3	80
10	2	0	2	16	12	8	38	6	80
11	2	0	3	39	13	11	15	2	80
12	2	0	4	4	9	21	22	14	70
13	2	1	1	20	14	13	27	6	80
14	2	1	2	1	4	8	32	35	80
15	2	1	3	21	15	5	35	4	80
16	2	1	4	5	4	18	23	19	69

Modèle logit multinomial
astuce

$$\left\{\begin{array}{l}\text{lien logit}\\ \text{multinomiale}\end{array}\right. \Leftrightarrow \left\{\begin{array}{l}\text{lien log}\\ \text{poisson}\\ \text{conditionnellement à la marge ligne}\end{array}\right.$$

ex. : triticale :

Facteur ligne	Facteurs	5 catégories	
1 ⋮ 16		n_{ij}	$n_{i.}$

Modèle à profil constant : Y' = ligne + catégorie

Facteur ligne	Facteurs	5 catégories		
1		$\widehat{p}_1$	$\ldots$	$\widehat{p}_5$
⋮		$\widehat{p}_1$	$\ldots$	$\widehat{p}_5$
16			⋮	

$$\widehat{p}_i = \frac{n_{.j}}{n_{..}}$$

Modèle logit multinomial
astuce

Modèles étudiés

$$\begin{array}{llr} Y' = \text{ligne} + \text{catégorie} & + \text{Mbase} & .\ \text{catégorie} \\ & + \text{Kin} & .\ \text{catégorie} \\ & + \text{Mbase . Kin} & .\ \text{catégorie} \\ & + \ldots & .\ \text{catégorie} \end{array}$$

Modèle saturé

Y' = ligne + catégorie + ligne : catégorie

5 catégories			
$\widehat{p_{11}}$	$\widehat{p_{12}}$	$\ldots$	$\widehat{p_{15}}$
$\widehat{p_{21}}$	$\widehat{p_{22}}$	$\ldots$	$\widehat{p_{25}}$
	$\vdots$		

$$\widehat{p_{ij}} = \frac{n_{ij}}{n_{i.}}$$

$$\begin{cases} \mathrm{F}(p_r) = \text{Mbase+Kin+Mbase.Kin}, r = 2\ldots5 \\ \mathrm{F} : \text{logit} \end{cases}$$

$$\Longleftrightarrow$$

$$\begin{cases} \mathrm{F}(n_{ij}) = \text{ligne+cat+Mbase.cat+Kin.cat+Mbase.Kin.cat} \\ \mathrm{F} : \log \end{cases}$$

Nouvelle présentation des données

exemple triticale

ligne	Mbase	Kin	Géno	Catégorie	$Y' = n_{ij}$
1	1	0	1	plante	20
2	1	0	2	plante	16
3	1	0	3	plante	38
4	1	0	4	plante	10
5	1	1	1	plante	16
6	1	1	2	plante	3
7	1	1	3	plante	13
8	1	1	4	plante	3
9	2	0	1	plante	21
...	...	...	...	...	...
16	2	1	4	plante	5
1	1	0	1	ébauche	10
...	...	...	...	...	...
16	2	1	4	ébauche	4
...	...	...	...	...	...
1	1	0	1	rien	5
...	...	...	...	...	...
16	2	1	4	rien	19

Modèle à profil constant (ou modèle NUL)

exemple triticale

Y' =
ligne + catégorie

	Df	Dev	Res.Df	Res.Dev	Pr.Chi
NULL			79	630.19	
ligne	15	4.17	64	626.01	0.997
cat	4	274.99	60	351.02	0.000

probabilités estimées

plante	ébauche	racine	cal	rien
0.20	0.12	0.12	0.39	0.17
0.20	0.12	0.12	0.39	0.17
0.20	0.12	0.12	0.39	0.17
		...		
0.20	0.12	0.12	0.39	0.17

Modèle additif

exemple triticale

Y' =
ligne + cat +
Mbase.cat + Kin.cat + Géno.cat

	Df	Dev	Res.Df	Res.Dev	Pr.Chi
NULL			79	630.19	
ligne	15	4.17	64	626.01	0.997
cat	4	274.99	60	351.02	0.000
Mbase.cat	4	12.78	56	338.24	0.012
Kin.cat	4	71.26	52	266.98	0.000
Géno.cat	12	176.04	40	90.94	0.000

Modèle interactif d'ordre 2

exemple triticale

Y' =
ligne + cat +
Mbase.cat + Kin.cat + Géno.cat +
Mbase.Kin.cat+Mbase.Géno.cat+Kin.Géno.cat

	Df	Dev	Res.Df	Res.Dev	Pr.Chi
NULL			79	630.19	
ligne	15	4.17	64	626.02	0.997
cat	4	274.99	60	351.02	0.000
Mbase.cat	4	12.78	56	338.24	0.012
Kin.cat	4	71.26	52	266.98	0.000
Géno.cat	12	176.04	40	90.93	0.000
Mbase.Kin.cat	4	1.76	36	89.18	0.780
Mbase.Géno.cat	12	17.25	24	71.92	0.140
Kin.Géno.cat	12	53.75	12	18.17	0.000

Modèle avec un seul terme d'interaction (modèle retenu)

exemple triticale

Y' =
ligne + cat +
Mbase.cat + Kin.cat + Géno.cat +
Kin.Géno.cat

	Df	Dev	Res.Df	Res.Dev	Pr.Chi
NULL			79	630.19	
ligne	15	4.17	64	626.02	0.997
cat	4	274.99	60	351.02	0.000
Mbase.cat	4	12.78	56	338.24	0.012
Kin.cat	4	71.26	52	266.98	0.000
Géno.cat	12	176.04	40	90.93	0.000
Kin.Géno.cat	12	53.29	28	37.64	0.000

Modèle à profil constant (ou modèle NUL)

exemple triticale

Y' =
ligne + catégorie

Source	Deviance	DF	ChiSquare	Pr>Chi
INTERCEPT	630.1957	0	.	.
ligne	626.0194	15	4.1763	0.9971
cat	351.0251	4	274.9942	0.0001

probabilité estimées

plante	ébauche	racine	cal	rien
0.20	0.12	0.12	0.39	0.17
0.20	0.12	0.12	0.39	0.17
0.20	0.12	0.12	0.39	0.17
		...		
0.20	0.12	0.12	0.39	0.17

Modèle additif

exemple triticale

Y' =
ligne + cat +
Mbase(cat) + Kin(cat) + Géno(cat)

Source	Deviance	DF	ChiSquare	Pr>Chi
INTERCEPT	630.1957	0	.	.
ligne	626.0194	15	4.1763	0.9971
cat	351.0251	4	274.9942	0.0001
Mbase(cat)	338.2405	4	12.7846	0.0124
Kin(cat)	266.9794	4	71.2611	0.0001
Géno(cat)	90.9367	12	176.0427	0.0001

Modèle interactif d'ordre 2

exemple triticale

Y' =
ligne + cat +
Mbase(cat) + Kin(cat) + Géno(cat) +
Mbase*Kin(cat) + Mbase*Géno(cat) +
Kin*Géno(cat)

Source	Deviance	DF	ChiSquare	Pr>Chi
INTERCEPT	630.1957	0	.	.
ligne	626.0194	15	4.1763	0.9971
cat	351.0251	4	274.9942	0.0001
Mbase(cat)	338.2405	4	12.7846	0.0124
Kin(cat)	266.9794	4	71.2611	0.0001
Géno(cat)	90.9367	12	176.0427	0.0001
Mbase*Kin(cat)	89.1769	4	1.7598	0.7798
Mbase*Géno(cat)	71.9250	12	17.2519	0.1404
Kin*Géno(cat)	18.1753	12	53.7497	0.0001

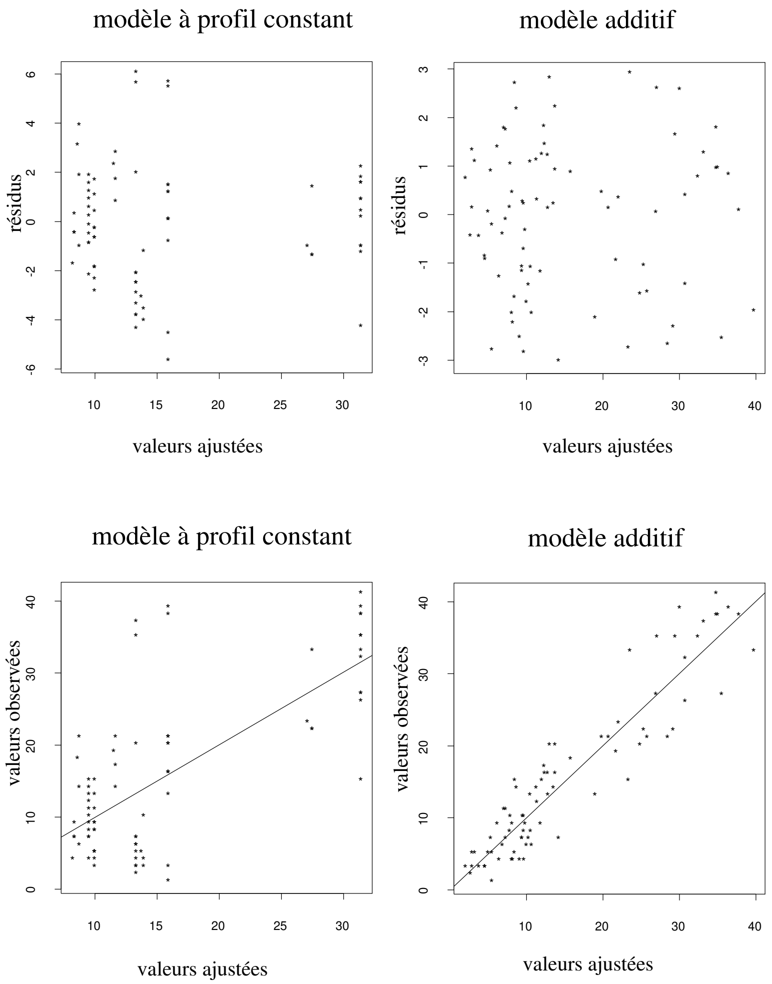
Modèle avec un seul terme d'interaction (modèle retenu)

exemple triticale

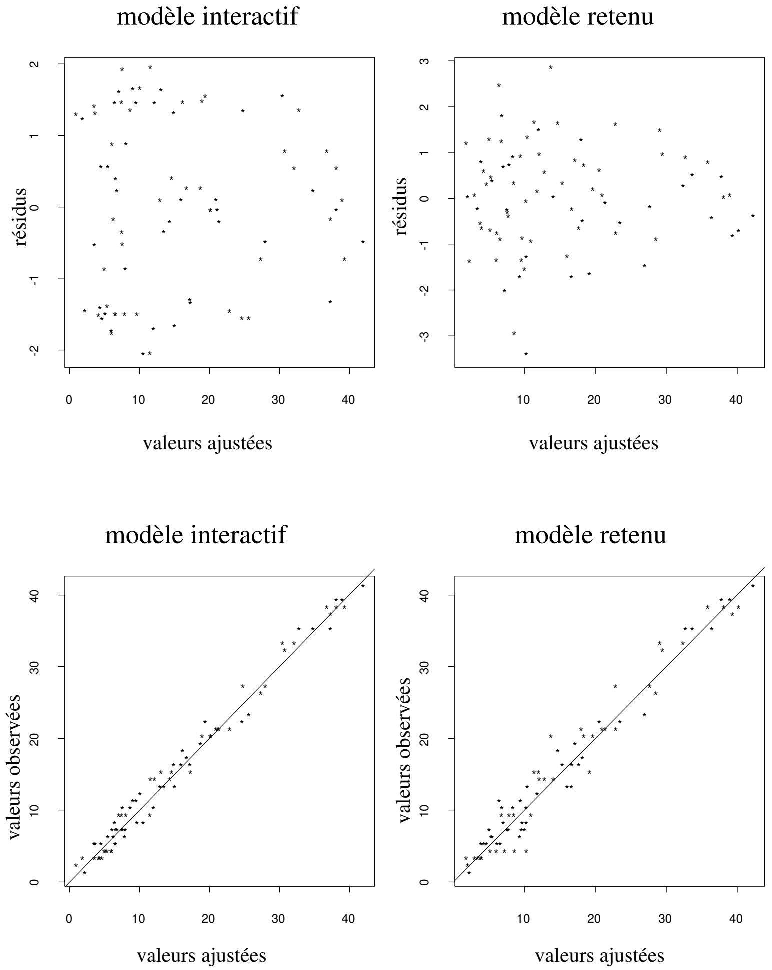
Y' =
ligne + cat +
Mbase(cat) + Kin(cat) + Géno(cat)
Kin*Géno(cat)

Source	Deviance	DF	ChiSquare	Pr>Chi
INTERCEPT	630.1957	0	.	.
ligne	626.0194	15	4.1763	0.9971
cat	351.0251	4	274.9942	0.0001
Mbase(cat)	338.2405	4	12.7846	0.0124
Kin(cat)	266.9794	4	71.2611	0.0001
Géno(cat)	90.9367	12	176.0427	0.0001
Kin*Géno(cat)	37.6440	12	53.2927	0.0001

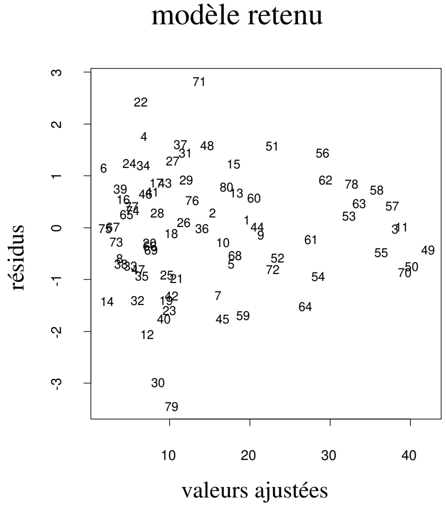
Modèle à profil constant, modèle additif

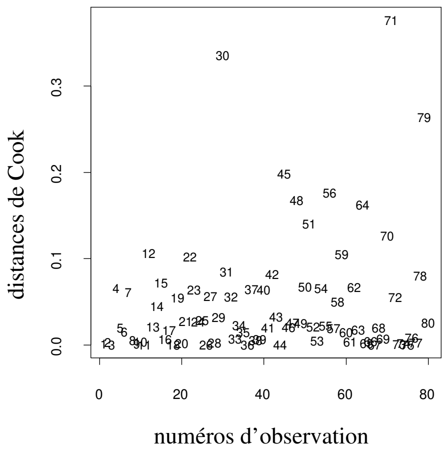


Modèle interactif d'ordre 2, modèle retenu



Modèle retenu - validation





Analyse génotype par génotype

génotype 1 :

	Df	Dev	Res.Df	Res.Dev	Pr.Chi
ligne	3	0.00	16	143.58	1.000
cat	4	127.81	12	15.76	0.000
cat.Mbase	4	1.65	8	14.11	0.799
cat.Kin	4	10.63	4	3.47	0.031

génotype 2 :

	Df	Dev	Res.Df	Res.Dev	Pr.Chi
ligne	3	0.00	16	200.762	1.000
cat	4	18.01	12	82.75	0.000
cat.Mbase	4	1.73	8	81.01	0.734
cat.Kin	4	74.96	4	6.05	0.000

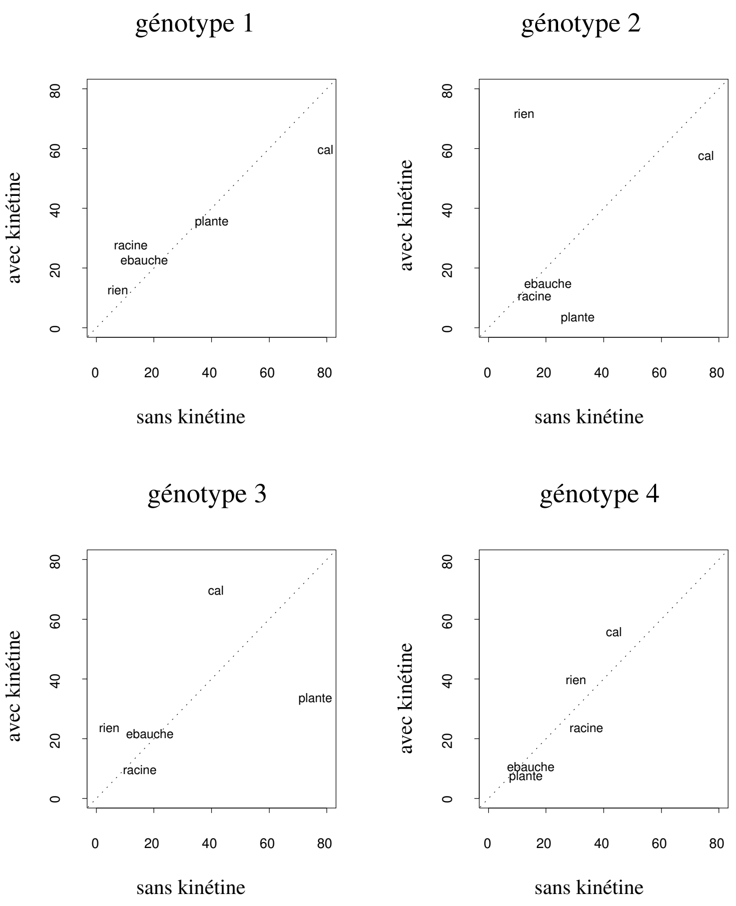
génotype 3 :

	Df	Dev	Res.Df	Res.Dev	Pr.Chi
ligne	3	0.00	16	183.01	1.000
cat	4	19.46	12	63.55	0.000
cat.Mbase	4	18.80	8	44.75	0.001
cat.Kin	4	40.46	4	4.28	0.000

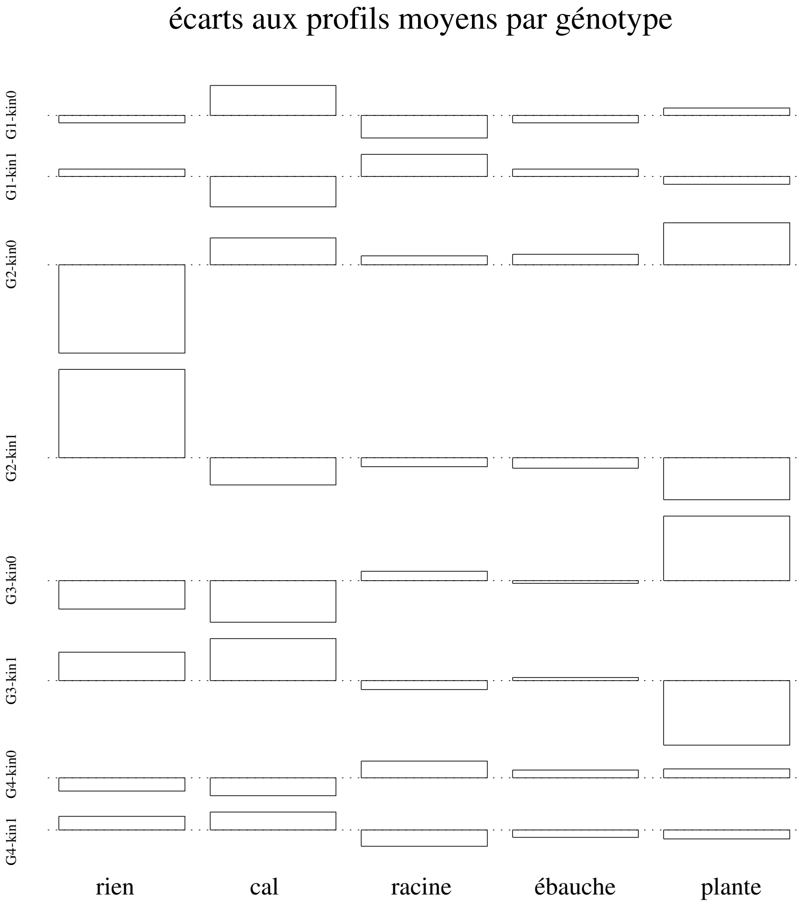
génotype 4 :

	Df	Dev	Res.Df	Res.Dev	Pr.Chi
ligne	3	0.01	16	98.67	1.000
cat	4	77.31	12	21.24	0.000
cat.Mbase	4	8.35	8	12.99	0.079
cat.Kin	4	7.27	4	5.73	0.120

Réponse des 4 génotypes à la kinétine



Réponse des 4 génotypes à la kinétine



Conclusion sur l'exemple triticale

AVANTAGE :

Par rapport à l'analyse catégorie par catégorie :

interprétation globale en terme de profil.

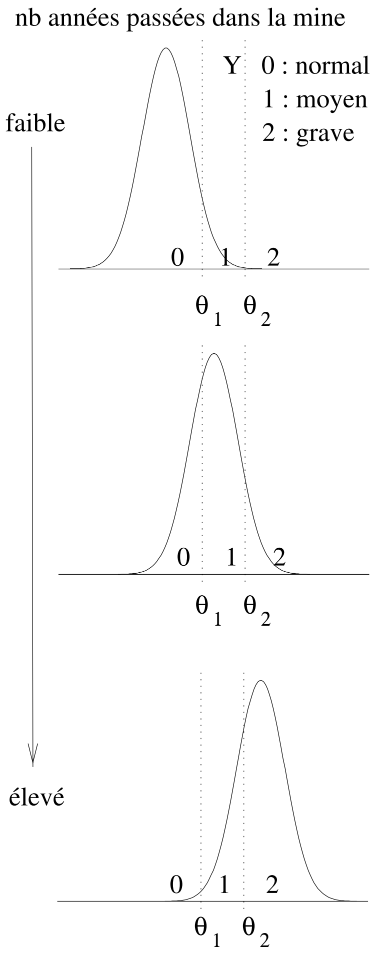
LIMITES :

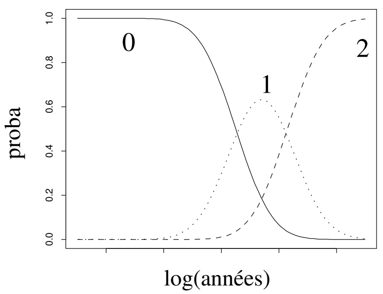
1. Nécessité de regrouper les catégories à effectif faible,
2. Ordre des catégories non pris en compte.

$\longrightarrow$ Modèles pour catégories ordonnées ?

catégories ordonnées

Approche seuil : seuils fixés





Catégories ordonnées

Approche seuil :

Modèle cumulatif simple (seuils fixés)

$$p_r = P(Y \leq r) = P(U \leq \theta_r) = F(\alpha_r + \beta x)$$

F: logistique $\Rightarrow$ modèle cumulatif logistique simple.

$$P(Y \leq r) = \frac{exp(\alpha_r + \beta x)}{1 + exp(\alpha_r + \beta x)}$$

$$\log\left(\frac{P(Y \leq r)}{P(Y > r)}\right) = \underbrace{\alpha_r}_{=\theta_r} + \beta\, x$$

contrainte $\theta_1 < \ldots < \theta_{q-1}$

proportional odds model

$\rightarrow$ ordre strict des populations

Catégories ordonnées

<u>Approche seuil</u> :

- modèle cumulatif : $P(Y \leq r) = F(\theta_r + \beta\ x)$
  séquentiel : $P(Y = r | Y \geq r) = F(\theta_r + \beta\ x)$
- modèle simple : seuils fixés
  généralisé : seuils variables

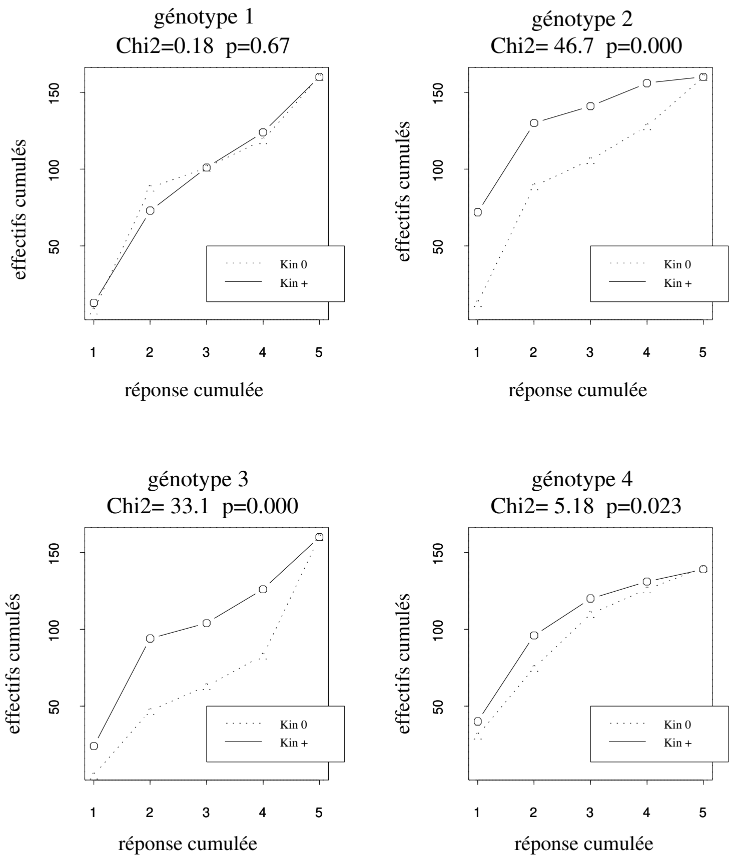
choix de F $\rightarrow$ logit
$\rightarrow$ extreme minimal value distribution
$\rightarrow$ extreme maximal value distribution
$\rightarrow$ ...

$\rightarrow$ $F^{-1}(\eta) = \log(\frac{p}{1-p})$ $F(p) = \frac{\exp \eta}{1+\exp \eta}$
$\rightarrow$ $F^{-1}(\eta) = \log(-\log(1-p))$ $F(p) = 1 - \exp(-\exp \eta)$
$\rightarrow$ $F^{-1}(\eta) = -\log(-\log p)$ $F(p) = \exp(-\exp(-\eta))$
$\rightarrow$ ...

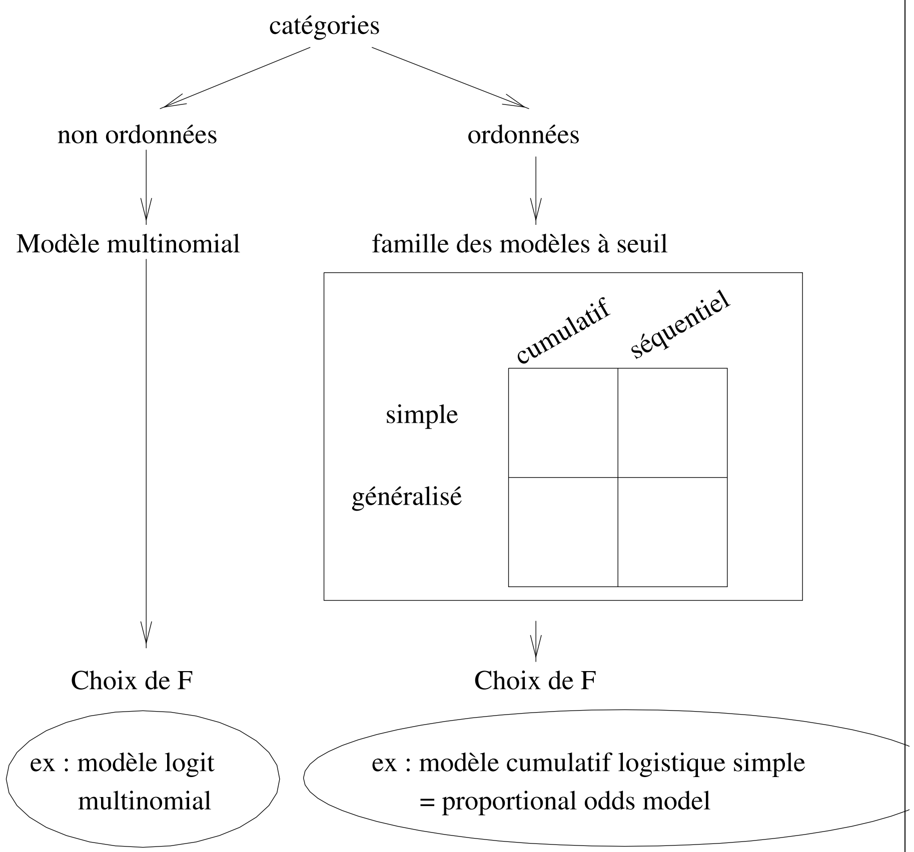
Ex. triticale, modèle cumulatif

Réponse cumulée des 4 génotypes à la kinétine

1=rien; 2=+cal; 3=+racines; 4=+ébauche; 5=+plante



RÉSUMÉ



catégories
non ordonnées
ordonnées
Modèle multinomial
famille des modèles à seuil
cumulatif
séquentiel
simple
généralisé
Choix de F
Choix de F
ex : modèle logit multinomial
ex : modèle cumulatif logistique simple = proportional odds model