

Échantillonnage à Différents Niveaux (individus, populations) et Modèles d'Intégration

Nigel Gilles YOCCOZ

Université de Tromsø

Norvège

Thème: Dynamique des Populations et Démographie

Quelles sont les Questions?

Quelques idées récentes sur:

Données: Qualité, Échantillonnage

Analyses: Modélisation...

Fonctions de vraisemblance “délicates”, nouvelles approches

Grand nombre de variables explicatives (p.e. climat)

Sélection de modèles

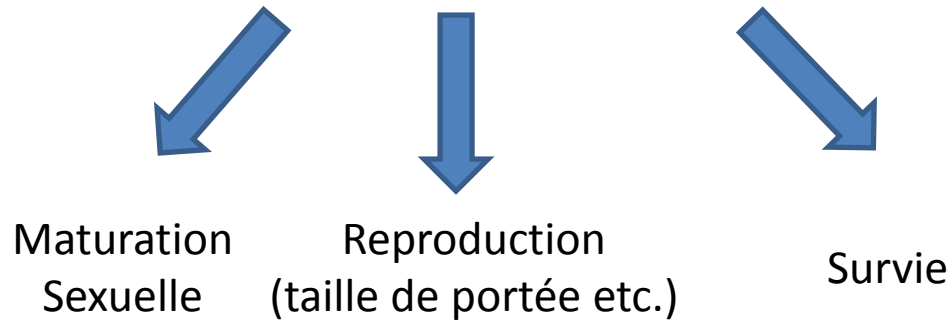
Mesure de l'effet

Mesure du support, de l'importance

(“evidence-based conservation”)

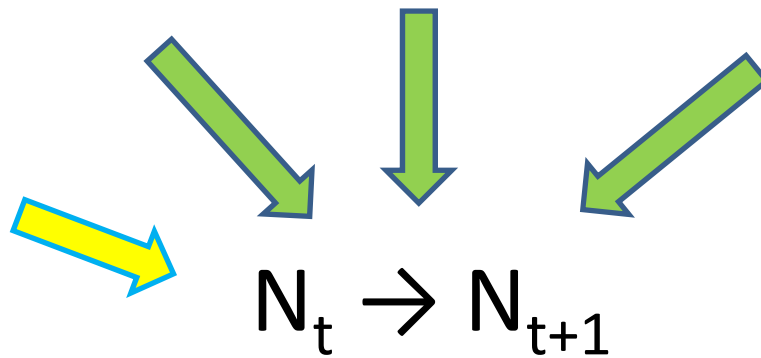
Dynamique des populations – Démographie

Causes multiples
Extrinsèques: Climat
Intrinsèques: Compétition [Densité]
(en général interagissent)



← Age
Poids ...

Dispersion (Im/Emigration)



Questions

Générales: Estimation des composantes de la dynamique, et caractérisation de cette dynamique (p.e. cyclique, chaotique)

Structure + Variabilité + Erreur de Mesure

Mécanismes: action du climat, importance relative de différentes variables, de différents paramètres démographiques

Gestion: Modèle prédictif: $g(N_{t+1} | N_t, \text{Action}_t)$

Statistical inference for noisy nonlinear ecological dynamic systems

Simon N. Wood¹

Modèle de Ricker

$$N_{t+1} = rN_t e^{-N_t + e_t}$$

Fonction de
Vraisemblance
Impossible à utiliser
directement

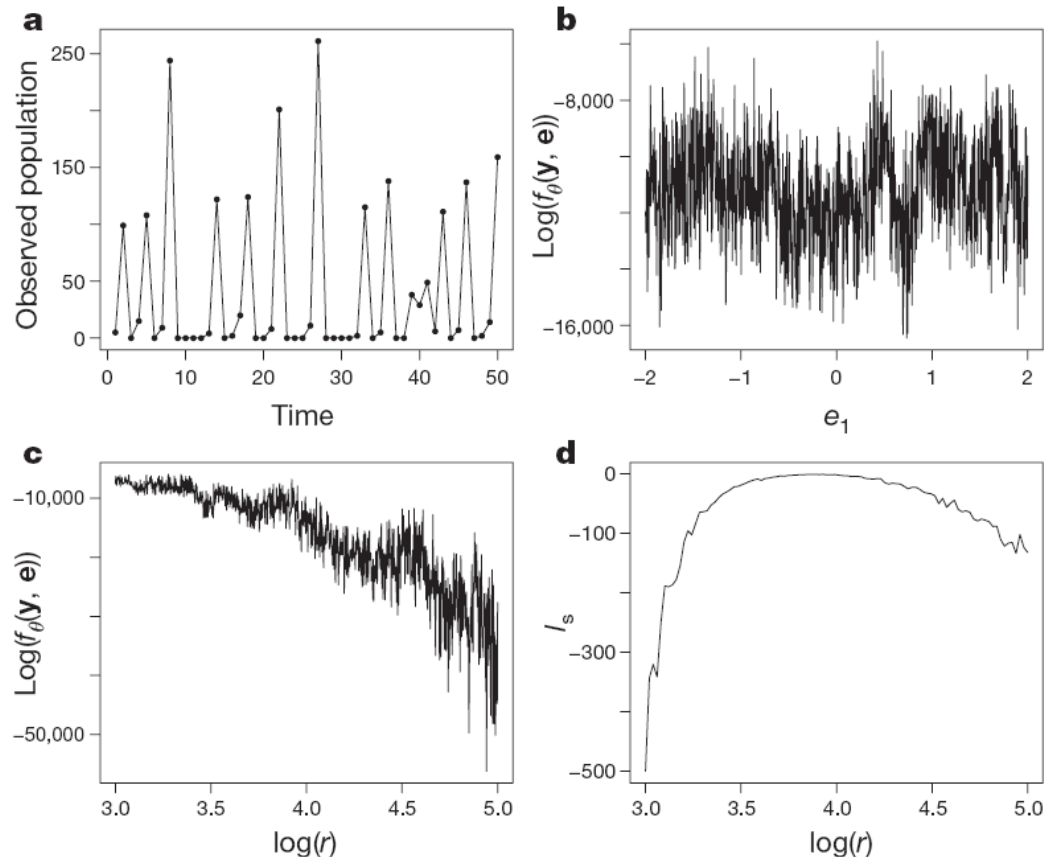
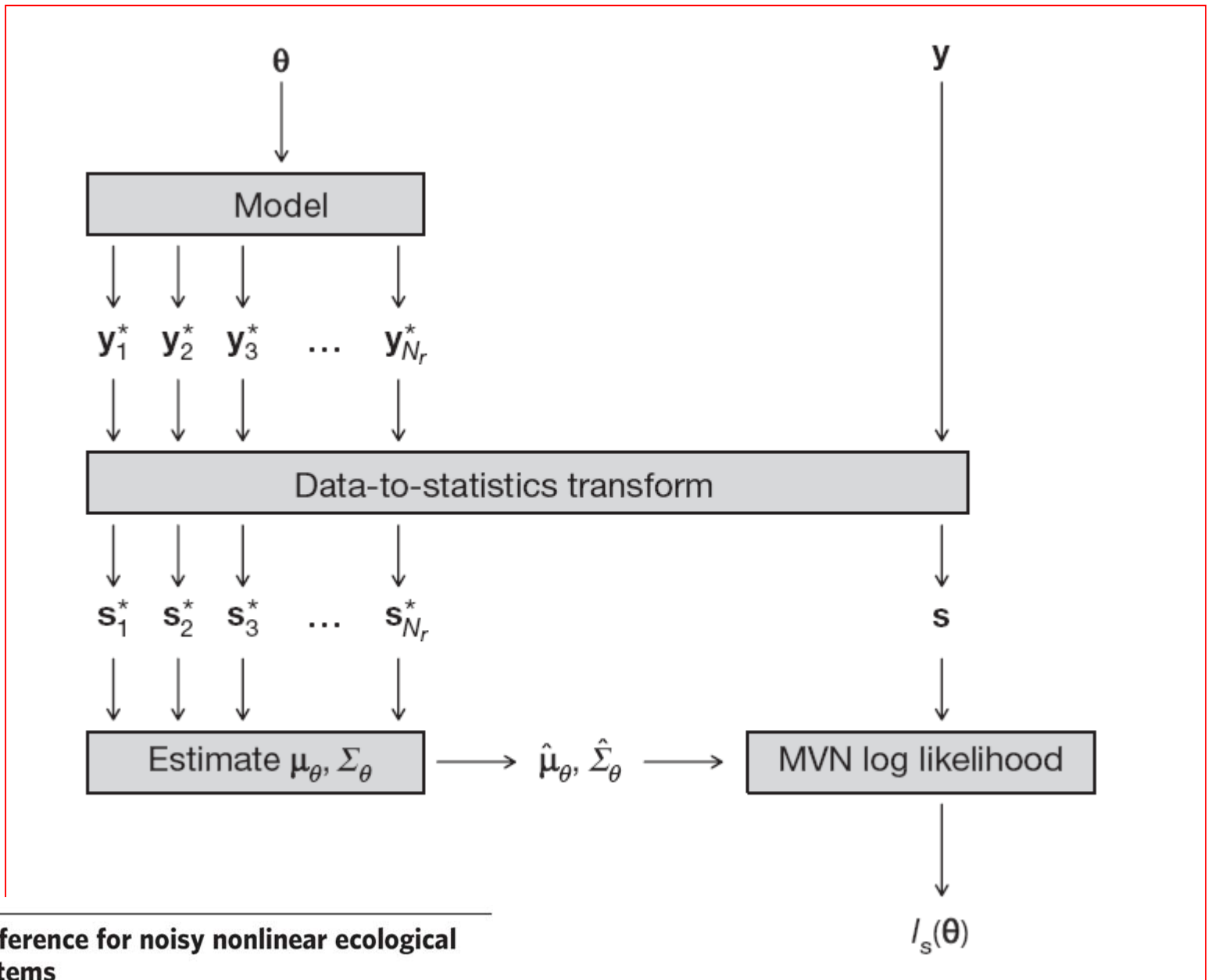


Figure 1 | Measuring fit of the Ricker model. **a**, Population data simulated from the Ricker model in the text, observed under Poisson sampling ($\log(r) = 3.8$, $\sigma = 0.3$, $\phi = 10$). **b**, The log joint probability density, $\text{log}(f_\theta(\mathbf{y}, \mathbf{e}))$, of data, \mathbf{y} , and random process noise terms, \mathbf{e} , plotted against the value of the first process noise deviate, e_1 , with the rest of \mathbf{e} and \mathbf{y} held fixed. **c**, $\text{Log}(f_\theta(\mathbf{y}, \mathbf{e}))$ plotted against model parameter r , again with \mathbf{e} and \mathbf{y} held fixed. **d**, The log synthetic likelihood, l_s , plotted against $\log(r)$ for the Ricker model and the data given in **a** ($N_r = 500$).



Statistical inference for noisy nonlinear ecological dynamic systems

Statistical inference for noisy nonlinear ecological dynamic systems

Simon N. Wood¹

Modèle de Ricker

$$N_{t+1} = rN_t e^{-N_t + e_t}$$

Fonction de
Vraisemblance
Impossible à utiliser
directement

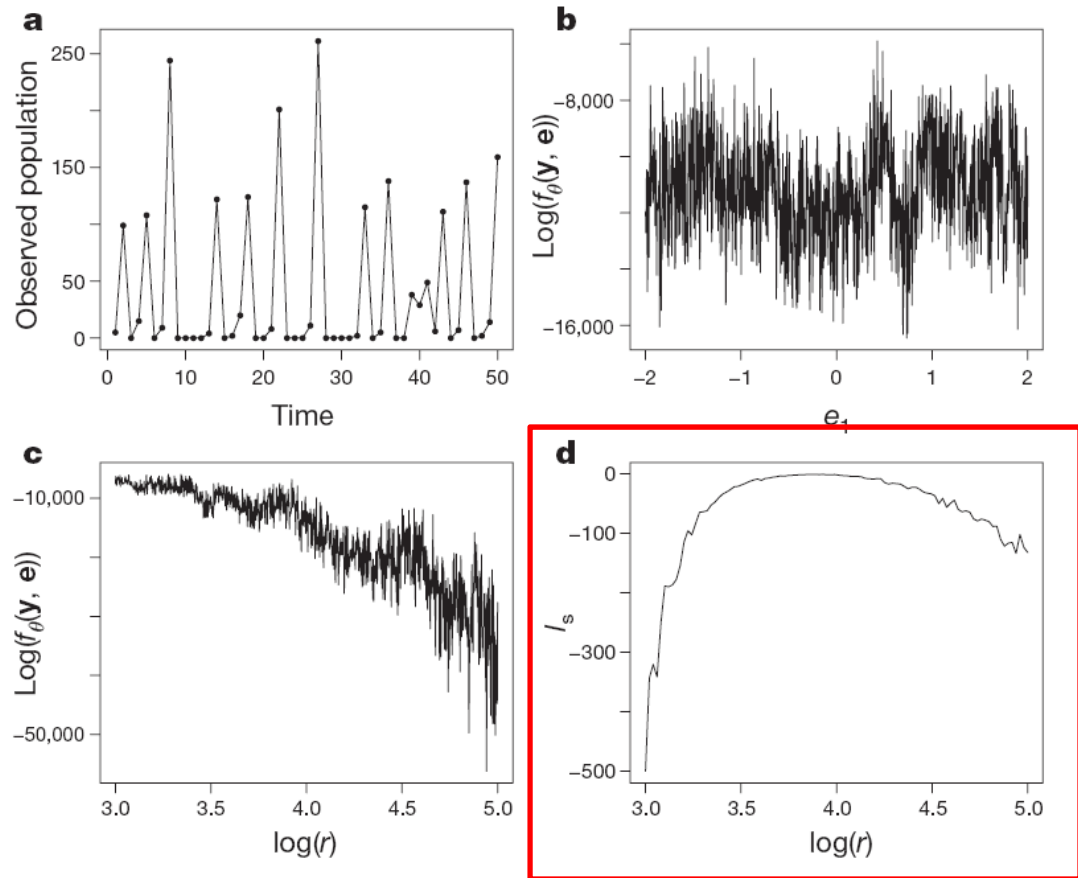


Figure 1 | Measuring fit of the Ricker model. **a**, Population data simulated from the Ricker model in the text, observed under Poisson sampling ($\log(r) = 3.8$, $\sigma = 0.3$, $\phi = 10$). **b**, The log joint probability density, $\text{log}(f_\theta(\mathbf{y}, \mathbf{e}))$, of data, \mathbf{y} , and random process noise terms, \mathbf{e} , plotted against the value of the first process noise deviate, e_1 , with the rest of \mathbf{e} and \mathbf{y} held fixed. **c**, $\text{Log}(f_\theta(\mathbf{y}, \mathbf{e}))$ plotted against model parameter r , again with \mathbf{e} and \mathbf{y} held fixed. **d**, The log synthetic likelihood, l_s , plotted against $\log(r)$ for the Ricker model and the data given in **a** ($N_r = 500$).

Recherches de mécanismes

- Multiples variables potentielles (p.e., descripteurs du climat)
- Différents paramètres démographiques (reproduction, survie, dispersion)
- Forte hétérogénéité
 - entre individus (âge, poids, sexe, ...)
 - entre populations (densité, ressources, pressions humaines)
- Difficultés à capturer tous les individus ou estimations des effectifs peu précises (et souvent biaisées)
- Modèles d'estimation pouvant avoir plusieurs centaines de paramètres, et combinant différentes informations (“integrated population models”)

Population de Chamois Modèle d'Intégration

Marquage de Cabris
peu après la naissance

Suivi par Observations
(revu – non revu)

Modèle Multinomial

Observations de Femelles
en été (avec/sans jeune)

- Jeune pas nécessairement vu
- probabilité d'observer une femelle peut dépendre de son statut
- Hétérogénéité (habitat)

GLMM (beta-binomial etc.)

Comptage en automne

N_t Effectif
↓
 C_t Compté

$$N_{t+1} = A_t N_t$$

$$C_t = \lambda_t N_t$$

Filtre de Kalman
"state-space model"

$$L_{\text{tot}}(Y, \theta) = L_c(Y_c, \theta_c) L_j(Y_j, \theta_j) L_N(Y_N, \theta_N)$$

Completing the Ecological Jigsaw

Panagiotis Besbeas, Rachel S. Borysiewicz, and Bryon J.T. Morgan

Les modèles devraient tenir compte de l'échantillonnage

Des individus (probabilité de capture)

Des populations (p.e., stratification variables climatiques)

“Models need to properly reflect features of the sample design such as weighting, stratification, and clustering or else inferences are likely to be distorted” (Little 2004)

Si les probabilités de sélection dépendent de X, elles devraient être incluses dans les modèles

$W = \text{Diag}(1/\pi_i)$ (probabilités de sélection)

$$\hat{\beta} = (X' X)^{-1} X' Y$$

$$V(\beta) = (X' X)^{-1} \sigma^2$$

$$\hat{\beta}_W = (X' W X)^{-1} X' W Y$$

$$V(\beta_W) = (X' W X)^{-1} (X' W^2 X) (X' W X) \sigma^2 \\ \neq (X' W X)^{-1} \sigma^2$$

La sélection de modèle reste un problème délicat

- AIC, BIC, etc.

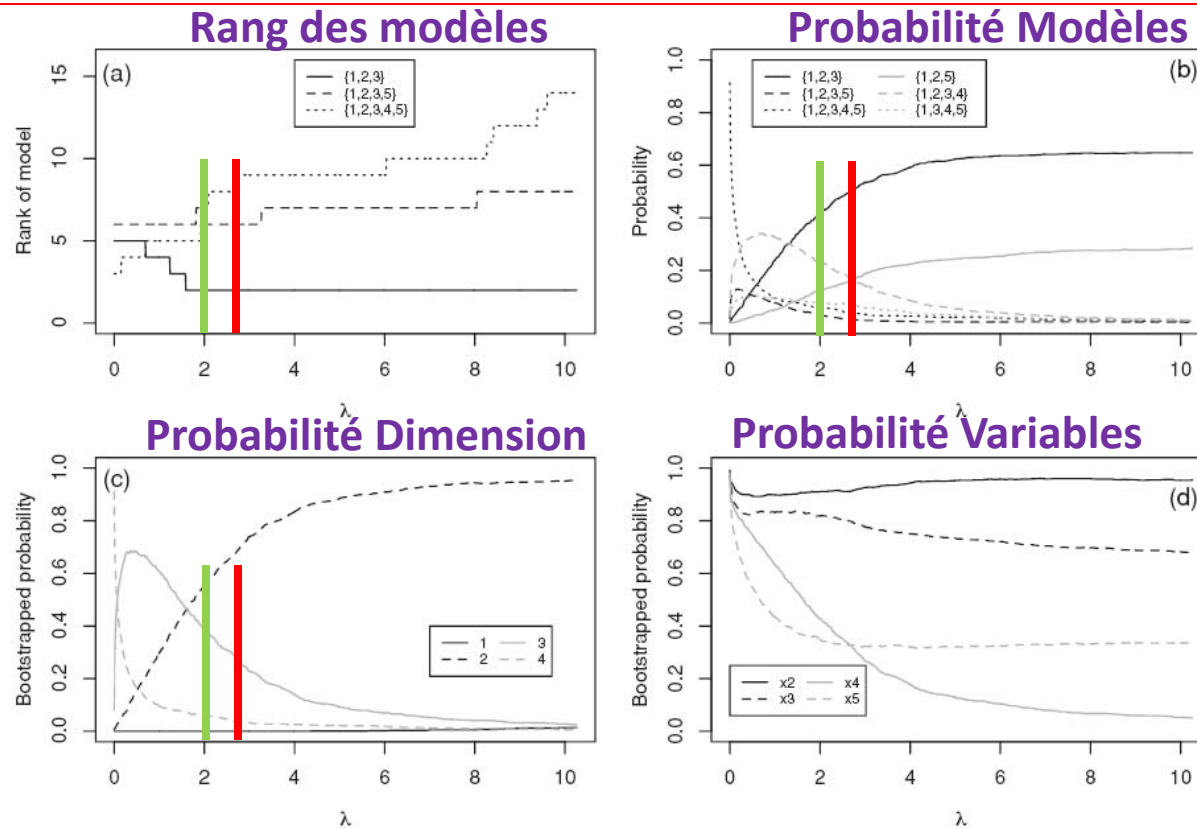
(description vs prediction)

Ajustement + Pénalité ($\sim \lambda \times p$)

International Statistical Review (2010), 78, 2, 240–256 doi:10.1111/j.1751-5823.2010.00108.x

On Model Selection Curves

Samuel Müller¹ and Alan H. Welsh²



N=13
Log(N)=2.6

Figure 2. Diagnostic plots for cement data: (a) rank plot; (b) single model plot; (c) dimensionality plot; (d) variable detection plot.

Sélection de Modèles (suite)

Identifier les variables importantes → effet supérieur à ε (pas 0)

Comment définir une échelle commune pour les effets ?

Mesures classiques en meta-analyse peu satisfaisantes (e.g. P-value, R^2)
“Most correlation coefficients should never be calculated” (Tukey, 1954)

Coefficients de régression ? mais ...

Standardisation (p.e. effet de la température et des précipitations)

$$\beta_T/sd(T) \text{ et } \beta_P/sd(P)$$

Mais $sd(T)$ et $sd(P)$ sont sensibles à l'échantillonnage ...

Quelles valeurs choisir?

Gestion (conservation, pops exploitées) basée sur le support quantitatif (evidence-based medicine; “Centre for Evidence-Based Conservation”)

Comment synthétiser l’information obtenue sur de nombreuses populations, à partir de méthodes différentes, avec des prédicteurs différents...

Méta-analyses, mais reprenant les données brutes

Fonctions de vraisemblance produit des fonctions de vraisemblance de chacune des études (Schweder; “Integrated Population Modelling”)

Comment mesurer l’importance relative des variables ?

Estimators of Relative Importance in Linear Regression Based on Variance Decomposition

The American Statistician, May 2007, Vol. 61, No. 2 139

Ulrike GRÖMPING

Variable Importance Assessment in Regression: Linear Regression versus Random Forest

Ulrike GRÖMPING

The American Statistician, November 2009, Vol. 63, No. 4

Hierarchical Partitioning (Lindeman et al, 1980)

Considérer toutes les permutations possibles des prédicteurs et faire la moyenne des contributions de chaque variable (en considérant les SCE séquentielles)

(Random Forest: permuter les valeurs d'une variable)

Très "ad hoc", propriétés pas très bien connues

Utilise des R^2 – comment développer un analogue pour les coefficients de régression?

Conclusions

- Complexité des modèles d'estimation (Nb paramètres, plusieurs composantes, ...)
- Sélection de Modèles (dépend des objectifs: description, estimation, prédiction)
- Echantillonnage jamais pris en compte
- Certains concepts sont encore très mal définis – “Importance”