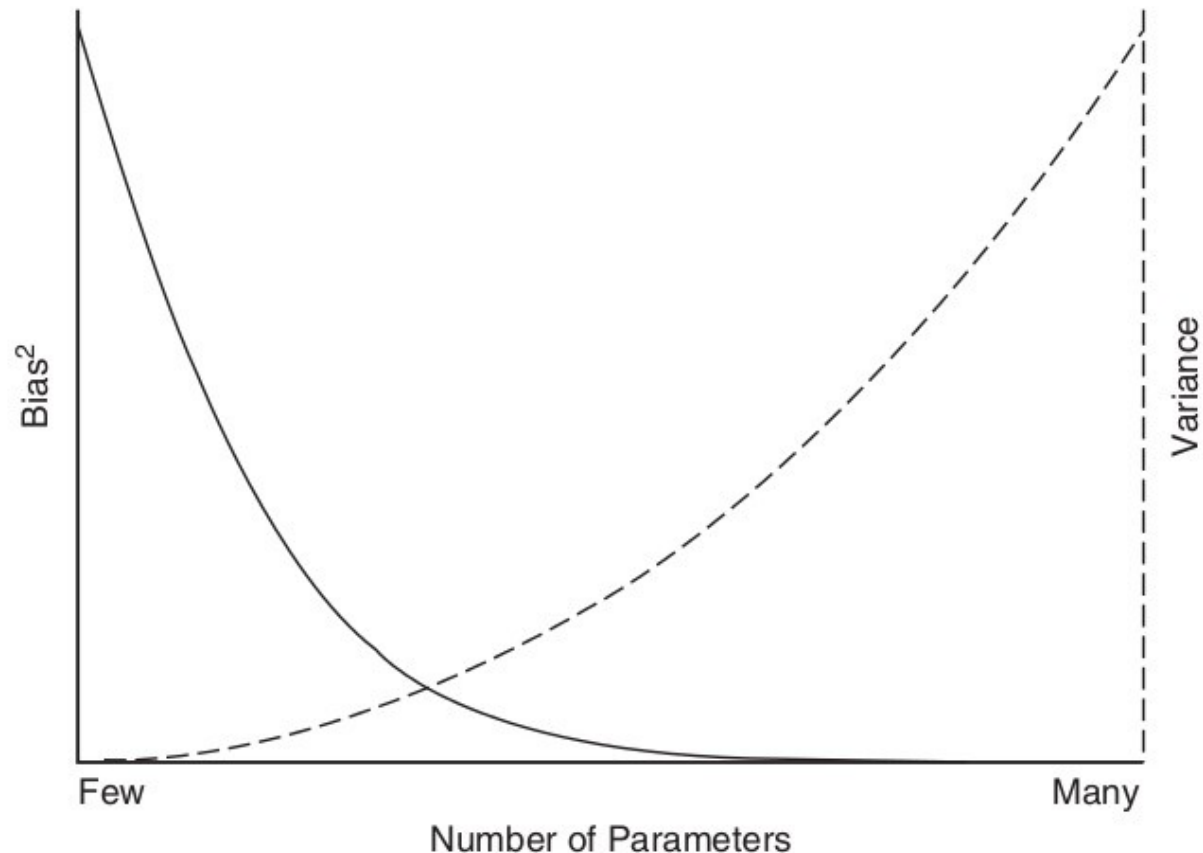


# Model selection and model averaging with an information criterion (AIC) approach



**G. San Martin**

gilles.sanmartin@gmail.com

Centre Wallon de Recherche Agronomique



# Outline

Example of ecological dataset/questions

Why do we need model selection ?

Basic use and theory about AIC (+ AICc, QAIC, QAICc)

Classical approach for model selection

Estimate model selection uncertainty and  
weight of evidence for each model

Classify the explanatory variables by order of importance

Inference from multiple models : model averaging

Some problems / questions ...

# Typical questions / datasets in ecology

where AIC methods are particularly useful

What is the weight of evidence of different competing hypothesis ?

Observative studies or experimental studies  
A few a priori competing hypothesis about a system

Which variables "explain" the best a system /  
are the most important for prediction ?

Generally observative - not experimental studies  
Complex system with many potential explanatory variables  
Ecologists want generally explanatory models not prediction only

# Typical questions / datasets in ecology

What is the weight of evidence of different competing hypothesis ?  
An example with predator-prey interactions



Asiatic ladybird



Other ladybird

*asymmetric*  
*"Intra-Guild Predation"*



Prey (aphids)

# Typical questions / datasets in ecology

What is the weight of evidence of different competing hypothesis ?

An example with predator-prey interactions

General hypothesis :

The Asiatic ladybirds will prefer eating aphids if possible and will eat native ladybirds (=IGP) only if they have no alternative

More precise hypothesis :

IGP increases when :

H1 : the number of aphids decreases

H2 : the number of native ladybirds increases

H3 : the total number of ladybirds increases relative to the aphids

# Typical questions / datasets in ecology

What is the weight of evidence of different competing hypothesis ?

An example with predator-prey interactions

In the field (~30 sites), measure :

Frequency of IGP

(=proportion of larvae with alkaloids from other species in their body)

= dependent variable

Abundance of aphids

Abundance of native ladybirds

Abundance of all ladybirds

# Typical questions / datasets in ecology

What is the weight of evidence of different competing hypothesis ?

An example with predator-prey interactions

→ translation of the different hypothesis in (binomial) models

```
mod1 <- IGP ~ native
```

```
mod2 <- IGP ~ aphids + ladybirdsTot,
```

```
mod3 <- IGP ~ native + aphids + native:aphids
```

```
mod4 <- IGP ~ aphids + ladybirdsTot + aphids:ladybirdsTot
```

```
mod5 <- IGP ~ native + aphids + ladybirdsTot + native:aphids + aphids:ladybirdsTot
```

```
mod6 <- IGP ~ native + aphids + ladybirdsTot
```

```
mod7 <- IGP ~ aphids + native
```

```
mod8 <- IGP ~ aphids
```

The models don't need to be nested or have similar additive structure.

You may want to compare directly mod1 and mod8

This is not possible with classical model comparison procedures  
(ie likelihood ratio test, F tests)

Model fit statistics (RMSE, RSS,  $R^2$ ) are unfair  
if the number of parameters is unequal

# Typical questions / datasets in ecology

Which variables "explain" the best a system /  
are the most important for prediction ?

Typical case in ecology :

Why is a species present or absent at a particular place ?

- Often : observative (not experimental) studies
- High level of complexity (many potential explanatory variables)

BUT

- Ecologists generally want explanatory models not prediction only



# Typical questions / datasets in ecology

Ex : which parameters explain the best the presence/absence of some bats species in subterranean cavities in winter ?

→ Prediction ie for protection purposes

→ Understanding ie : Why are there no bats in this kind of caves ?

## **Cavity characteristics :**

- length (log scale)
- height (log scale)
- presence of water
- humidity on the walls
  - temperature
  - hygrometry
- level of disturbance

## **Landscape around the cavity :**

- distance to the closest river
- number of other cavities
  - hedges
- urban areas
  - fields
- grassland
  - orchards
  - rivers
  - lakes
- seminatural biotopes

# Typical questions / datasets in ecology

Ex : which parameters explain the best the presence/absence of some bats species in subterranean cavities in winter ?

```
glm(formula = M.myotis ~ urban + fields + orchards + grasslands + deciduous +  
seminatural + rivers + lakes + nb_cavities + dist_river + lin_hedges + wall_humidity +  
water + logdevl + disturbance + logheight + tempmean + hygromean, family = binomial,  
data = data, subset = nbwinters > nbwinterslim)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-3.04187	0.99546	-3.056	0.00225	**
urban	-1.42223	0.62316	-2.282	0.02247	*
fields	-2.18103	0.87334	-2.497	0.01251	*
orchards	-0.22955	0.41618	-0.552	0.58124	
grasslands	0.32507	0.54315	0.598	0.54952	
deciduous	0.09538	0.56314	0.169	0.86551	
seminatural	-0.48839	0.47160	-1.036	0.30038	
rivers	0.01644	0.45267	0.036	0.97102	
lakes	-0.57397	1.61835	-0.355	0.72284	
nb_cavities	0.40285	0.29105	1.384	0.16632	
dist_river	0.54126	0.38111	1.420	0.15555	
lin_hedges	-0.28011	0.41052	-0.682	0.49503	
wall_humidity1	-0.15731	0.90433	-0.174	0.86190	
water1	1.79517	0.89592	2.004	0.04510	*
logdevl	0.30211	0.46528	0.649	0.51614	
disturbancedisturbed	2.82895	1.09409	2.586	0.00972	**
disturbancequiet	0.05182	0.87223	0.059	0.95263	
logheight	1.09343	0.41269	2.650	0.00806	**
tempmean	-0.20136	0.36082	-0.558	0.57679	
hygromean	0.95243	0.64822	1.469	0.14175	

# Why do we need model selection ?

"Finding the best approximating model"

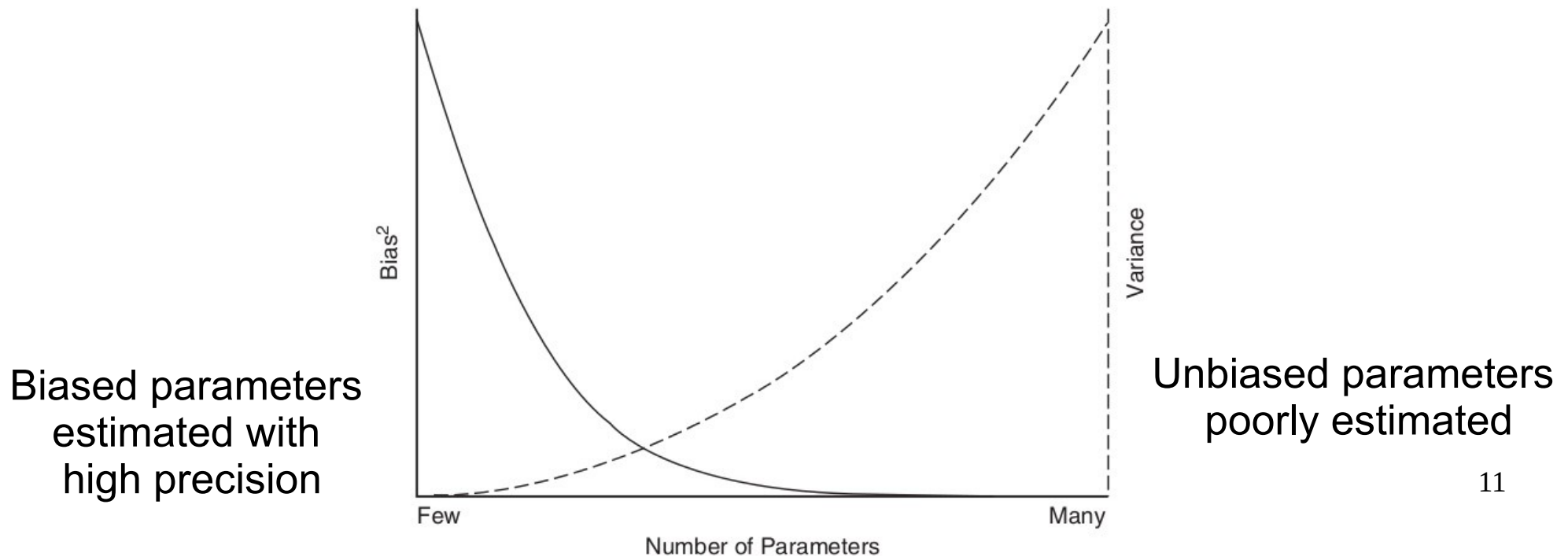
Trade-off : underfitting vs overfitting

Model that do not describe correctly the data

VS

Model that is too close to the data

(prediction cannot be generalized to the population)



# Why do we need model selection ?

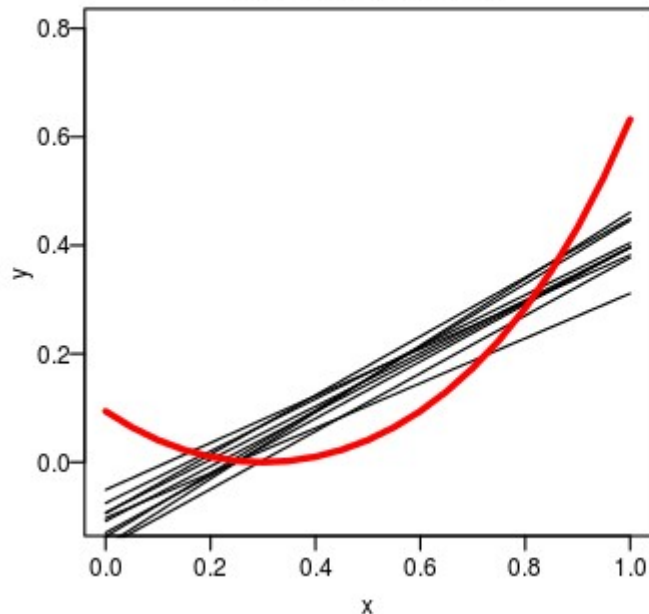
"Finding the best approximating model"  
Trade-off : underfitting vs overfitting

```
x <- seq(0, 1, 0.05)  
y <- exp((x-0.3)^2) - 1 + rnorm(21, 0, 0.1)
```

Too few parameters  
Biased but estimated with  
high precision

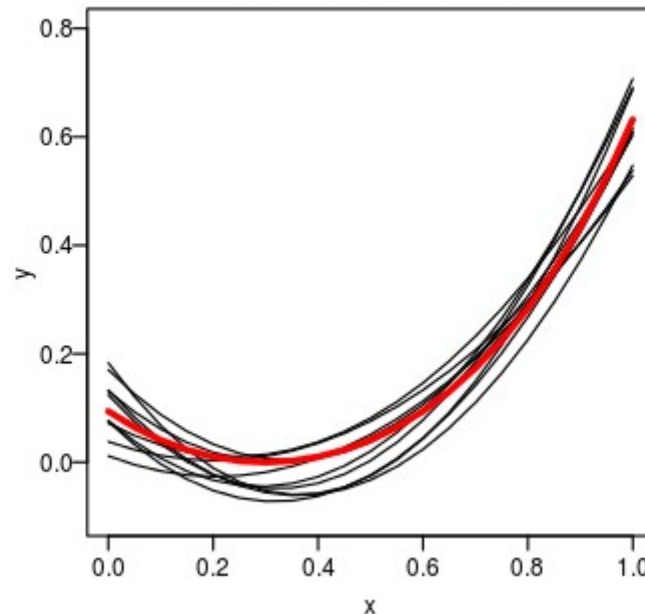
Too many parameters  
Unbiased but poorly estimated

1st order polynomial



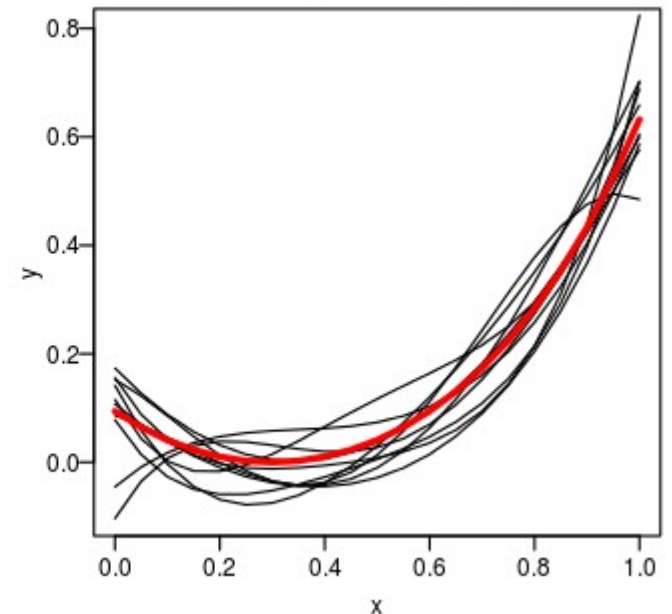
$$Y \sim x$$

2nd order polynomial



$$Y \sim x + x^2$$

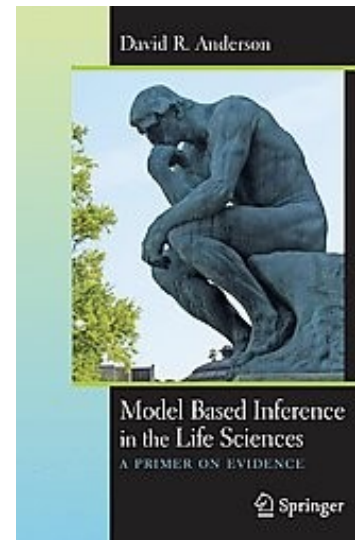
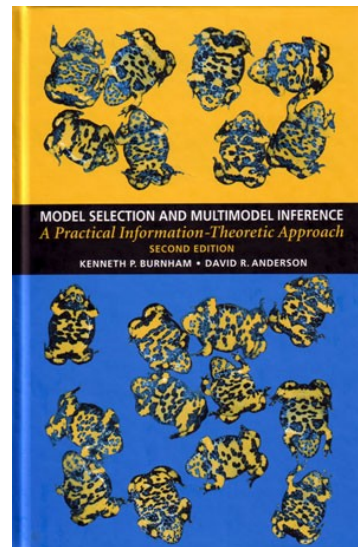
4th order polynomial



$$Y \sim x + x^2 + x^3 + x^4$$

# Information Criterion approach with AIC

Solutions presented in the books of Burnham & Anderson  
Classical AIC + much more



Development of tools for easy implementation of these methods :  
SAS (N. Schtickzelle - UCL) & R (myself : `model.select`)

# Information Criterion approach with AIC

Development of tools for easy implementation of these methods :  
SAS (N. Schtickzelle - UCL) & R (myself : `model.select`)

Today : at least 2 R packages to do this : `MuMin` + `AICcmodavg`

These packages are much more general  
(applicable to more models types)

`model.select` has been developed to be used with  
`lm`, `glm`, `lme` and `(g)lmer` models only.

It seems to be reliable for these few target cases  
It is more rapid (~3 times) and much less RAM demanding relative to  
`MuMin` (comparison made in 2011)

look here for a comparison (in french) of `MuMin` & `model.select`:  
<http://forums.cirad.fr/logiciel-R/viewtopic.php?t=3517>

## Basic AIC use and theory

**Kullback-Leibler distance** : quantity of information lost when you approximate full reality ( $f$ ) by a model ( $g$ )

**AIC (Akaike Information Criterion)** : asymptotically unbiased predictor of the relative expected K-L distance based on the maximized log-likelihood

$$\text{AIC} = -2 \log (\text{Lik} (\hat{\theta} | y)) + 2K$$



Maximized Likelihood                      Nbr of parameters

→ best approximating model  
= Model minimizing K-L distance  
= model with lowest AIC relative to other models

## Basic AIC use and theory

The AIC is also often presented as a measure of the fit quality of a model (likelihood) penalized by the complexity of the model (2K : number of parameters)

$$\text{AIC} = -2 \log ( \text{Lik} (\text{theta.hat} | y)) + 2K$$



NB : Likelihood =  
probability of the parameters theta.hat  
knowing the data y



Nbr of parameters  
= complexity



# Basic AIC use and theory

Difference of philosophy between **AIC** and **BIC**  
(Bayesian Information Criterion)

$$\mathbf{AIC} = -2 \text{ LogLik} + 2K$$

Full reality is a model with an infinite number of explanatory variables  
with tapering effects

You want to estimate how much information you lose when you  
approximate this with a finite model

$$\mathbf{BIC} = -2 \text{ LogLik} + K \log(n)$$

Full reality is a model with a finite number of explanatory variables  
You look for this "true model" within the set of models you are  
comparing

- more penalty on the number of parameters
- smaller more parsimonious models

# Basic AIC use and theory

## Other AIC related criterion

Standard AIC

$$\text{AIC} = -2 \log (\text{MaxLikelihood}) + 2K$$

**AIC for small sample size** (n) relative to the nbr of parameters (K)

$$\text{AICc} = \text{AIC} + (2K(K+1) / (n-K-1))$$

—► Always use this one

## AIC for overdispersed models

$$\text{QAIC} = -2 \log (\text{MaxLikelihood} / \text{c.hat}) + 2K$$

$$\text{QAICc} = \text{QAIC} + (2K(K+1) / (n-K-1))$$

c.hat is the overdispersion parameter  
estimated by the residual deviance / residual df  
or via Pearson residuals

**K = nbr of parameters + 1 !**

# Basic AIC use and theory

In R (Q)AICc generally not provided but easy to compute

```
aic <- function (model) {
  smry <- summary(model)
  k <- attr(logLik(model), "df")
  c <- overdisp(model)[1] # based on Pearson residuals
  if(length(c)!=0) {if((c)<1 ) c<-1}
  n <- length(fitted(model))
  loglik <- as.numeric(logLik(model))
  AIC <- (-2*loglik)+ (2*k)
  AICc <- (-2*loglik)+ (2*k) + (2*k*(k+1)/(n-k-1))
  k <- k+1
  QAIC <- -(2*loglik/c)+ (2*k)
  QAICc <- QAIC + (2*k*(k+1)/(n-k-1))
  (result <- cbind(AIC,AICc,QAIC,QAICc))
}

# With Overdispersion function :

overdisp <- function(mod) {
  k <- attr(logLik(mod), "df") # nb of parameters
  n <- length(fitted(mod)) # nb of observations
  pearsonresid <- (1/(n-k)) * sum(resid(mod, "pearson")^2)
  dev <- deviance(mod)/(n-k)
  result <- c(pearsonresid, dev)
  names(result) <- c("pearsonresid", "deviance")
  return(result)
}
```

# Basic AIC use and theory

AICc and QAICc allow you to order the models from the "best" to the "worse" (in K-L sense)

→ **only valid relative to the set of considered models**

If all models are bad, you will just find the best of the bad models

Do not use models without support from biological knowledge

Include a model with just the intercept

Check the final model fit

→ **only valid relative to a given dataset**

The y must remain exactly the same between the models

Be careful with NA values in the x

Do not mix y with different scales : eg y, log(y) etc...

When n increases, AIC and AICc values converge

→ always use the small sample versions AICc and QAICc

# Classical approach to model selection

Two approaches are frequently used for model selection/comparison

1) Sequential Null hypothesis testing

2) Sequential AIC comparisons

= Backward / forward / stepwise selection  
based on a null hypothesis criterion or an AIC criterion

# Classical approach to model selection

An example with Birthwt dataset from MASS package

Under-weighted newborn childs (yes/no)

vs mother characteristics :

age, weight (lwt), race, smoking habits,  
number of previous premature labours (ptl),  
hypertention (ht), uterine irritability (ui),  
number of physician visits during the first trimester (ftv).

```
library(MASS)
data(birthwt)

# tranform some variables into factors
birthwt[,c("race", "smoke", "ht", "ui")] <-
  lapply(birthwt[,c("race", "smoke", "ht", "ui")], factor)
birthwt$ptd <- factor(birthwt$ptl > 0)
```

# Classical approach to model selection

```
> mod <- glm(low ~ (age + lwt + race + smoke + ptd + ht + ui + ftv),  
+ family=binomial, data=birthwt)  
> summary(mod)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	0.644476	1.223889	0.527	0.59849	
age	-0.039548	0.038305	-1.032	0.30186	
lwt	-0.015078	0.007034	-2.143	0.03207	*
race2	1.218791	0.533168	2.286	0.02226	*
race3	0.819439	0.450466	1.819	0.06890	.
smoke1	0.859459	0.409836	2.097	0.03599	*
ptdTRUE	1.218512	0.463015	2.632	0.00850	**
ht1	1.860429	0.708161	2.627	0.00861	**
ui1	0.719299	0.463419	1.552	0.12062	
ftv	0.050900	0.175456	0.290	0.77174	

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 234.67 on 188 degrees of freedom  
Residual deviance: 196.75 on 179 degrees of freedom  
AIC: 216.75
```


# Classical approach to model selection

## Model selection with Null Hypothesis testing

```
> mod <- glm(low ~ (age + lwt + race + smoke + ptd + ht + ui + ftv),  
             family=binomial, data=birthwt)  
> drop1(mod, test="Chisq")
```

	Df	Deviance	AIC	LRT	Pr(Chi)	
<none>		196.75	216.75			
age	1	197.84	215.84	1.0878	0.296958	
lwt	1	201.83	219.83	5.0832	0.024158	*
race	2	203.24	219.24	6.4847	0.039071	*
smoke	1	201.25	219.25	4.4964	0.033966	*
ptd	1	203.83	221.83	7.0809	0.007791	**
ht	1	204.01	222.01	7.2628	0.007040	**
ui	1	199.12	217.12	2.3660	0.124006	
ftv	1	196.83	214.83	0.0836	0.772449	

We remove the "less significant"  
term for the next step



```
> mod <- glm(low ~ (age + lwt + race + smoke + ptd + ht + ui)  
             ,family=binomial, data=birthwt)  
> drop1(mod, test="Chisq")
```

	Df	Deviance	AIC	LRT	Pr(Chi)	
<none>		196.83	214.83			
age	1	197.85	213.85	1.0179	0.313026	
lwt	1	201.83	217.83	4.9996	0.025353	*
race	2	203.24	217.24	6.4066	0.040628	*
smoke	1	201.25	217.25	4.4133	0.035659	*
ptd	1	203.95	219.95	7.1144	0.007647	**
ht	1	204.01	220.01	7.1793	0.007375	**
ui	1	199.15	215.15	2.3177	0.127909	



# Classical approach to model selection

## Model selection with Null Hypothesis testing

```
> mod <- glm(low ~ (lwt + race + smoke + ptd + ht + ui),family=binomial,  
data=birthwt)  
> drop1(mod, test="Chisq")
```

	Df	Deviance	AIC	LRT	Pr (Chi)	
<none>		197.85	213.85			
lwt	1	203.82	217.82	5.9643	0.014599	*
race	2	205.47	217.47	7.6142	0.022212	*
smoke	1	202.57	216.57	4.7150	0.029900	*
ptd	1	204.22	218.22	6.3651	0.011639	*
ht	1	205.16	219.16	7.3106	0.006855	**
ui	1	200.48	214.48	2.6307	0.104817	

```
> mod <- glm(low ~ (lwt + race + smoke + ptd + ht),family=binomial,  
data=birthwt)  
> drop1(mod, test="Chisq")
```

	Df	Deviance	AIC	LRT	Pr (Chi)	
<none>		200.48	214.48			
lwt	1	207.16	219.16	6.6824	0.009737	**
race	2	207.96	217.96	7.4731	0.023836	*
smoke	1	205.40	217.40	4.9149	0.026626	*
ptd	1	208.25	220.25	7.7652	0.005326	**
ht	1	207.04	219.04	6.5572	0.010446	*

Final model with only significant terms

# Classical approach to model selection

## Problems :

- Final model dependent on the method (backward, forward,...)
- Multiple testing without p-value correction + arbitrary alpha level
- Removing large coefficients because of their large standard error can be "bad" (increasing bias)
  - ML estimates and standard errors valid only given a model (conditional standard error)
    - ie : standard error of the final model underestimated because they do not take into account model selection process

→

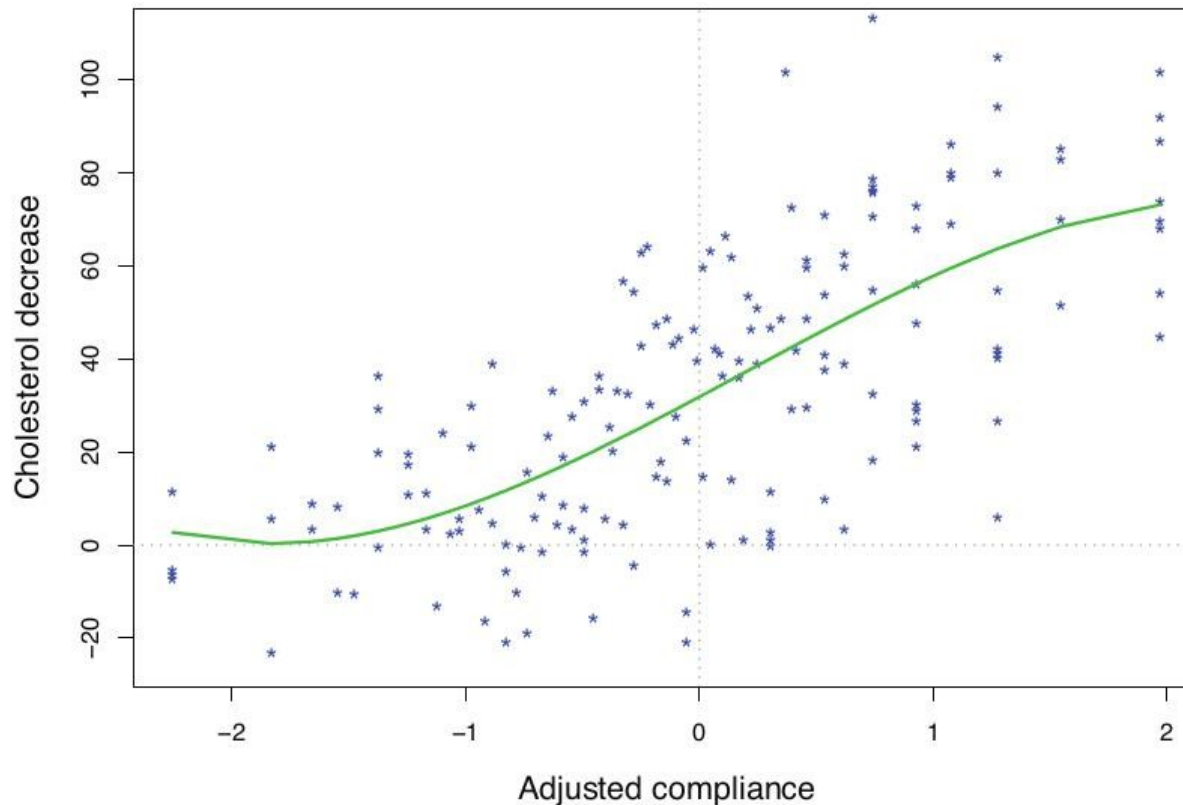
Generally strongly discouraged method

Generally OK if you test each effect once

(ie in experimental designs with few explanatory variables relative to the number of data)

# Classical approach to model selection

## Inference after model selection



Best polynomial degree = 3  
Determined here with Mallows  
 $C_p$  (similar use as AIC)

Degree	$C_p$
0	71887
1	1132
2	1412
3	667
4	1591
5	1811
6	2758

**Figure 20.1** Cholesterol data: cholesterol decrease plotted versus adjusted compliance for 164 men taking **cholestyramine**. The green curve is OLS cubic regression, with “cubic” selected by the  $C_p$  criterion. How accurate is the fitted curve?

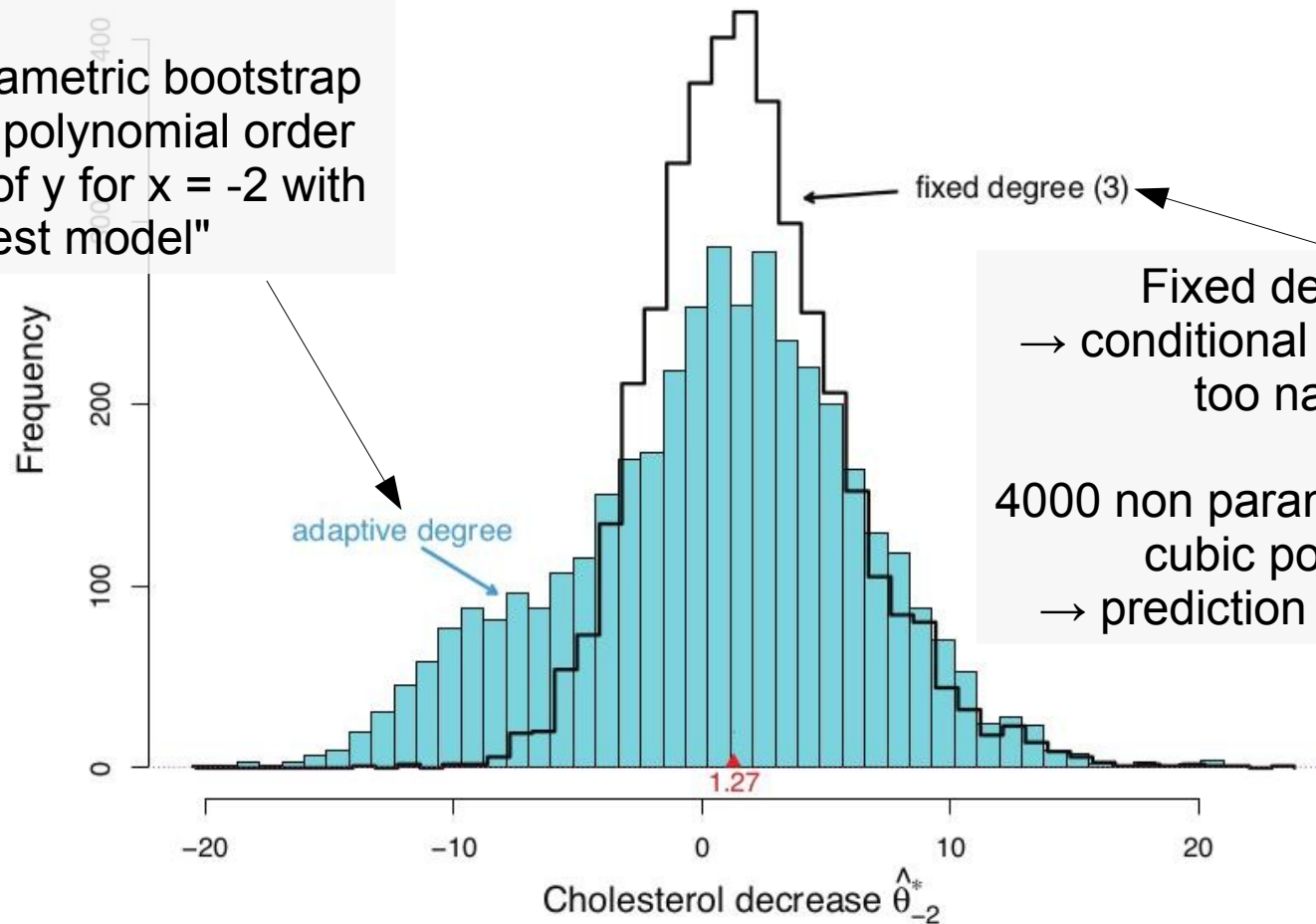
# Classical approach to model selection

## Inference after model selection

"Adaptive degree"

→ unconditional standard error broader !

4000 non parametric bootstrap  
Select "best" polynomial order  
→ prediction of  $y$  for  $x = -2$  with the "best model"



fixed degree (3)

Fixed degree = 3

→ conditional standard error too narrow !

4000 non parametric bootstrap  
cubic polynomial  
→ prediction of  $y$  for  $x = -2$

# Classical approach to model selection

## Model selection with stepwise AIC procedure

```
> library(MASS)
> stepAIC(mod, direction= "backward")
Start:  AIC=216.75
low ~ (age + lwt + race + smoke + ptd + ht + ui + ftv)
```

	Df	Deviance	AIC
- ftv	1	196.83	214.83
- age	1	197.84	215.84
<none>		196.75	216.75
- ui	1	199.12	217.12
- race	2	203.24	219.24
- smoke	1	201.25	219.25
- lwt	1	201.83	219.83
- ptd	1	203.83	221.83
- ht	1	204.01	222.01

If you remove one of these 2 explanatory variables from the model the AIC is lower relative to the full model with all explanatory variables <none>

```
Step:  AIC=214.83
low ~ age + lwt + race + smoke + ptd + ht + ui
```

	Df	Deviance	AIC
- age	1	197.85	213.85
<none>		196.83	214.83
- ui	1	199.15	215.15
- race	2	203.24	217.24
- smoke	1	201.25	217.25
- lwt	1	201.83	217.83
- ptd	1	203.95	219.95
- ht	1	204.01	220.01

# Classical approach to model selection

## Model selection with stepwise AIC procedure

Step: AIC=213.85

```
low ~ lwt + race + smoke + ptd + ht + ui
```

	Df	Deviance	AIC
<none>		197.85	213.85
- ui	1	200.48	214.48
- smoke	1	202.57	216.57
- race	2	205.47	217.47
- lwt	1	203.82	217.82
- ptd	1	204.22	218.22
- ht	1	205.16	219.16

If you remove any of the explanatory from this model, the AIC increases relative to the full model  
→ the algorithm stops

```
Call: glm(formula = low ~ lwt + race + smoke + ptd + ht + ui, family = binomial,  
data = birthwt)
```

Coefficients:

(Intercept)		lwt	race2	race3	smoke1	ptdTRUE
ht1	ui1					
-0.12533	-0.01592	1.30086	0.85441	0.86658	1.12886	
1.86690	0.75065					

# Classical approach to model selection

## Model selection with stepwise AIC procedure

```
> m1 <- glm(low ~ lwt + race + smoke + ptd + ht + ui,  
+ family=binomial, data=birthwt)  
> summary(m1)
```

Call:

```
glm(formula = low ~ lwt + race + smoke + ptd + ht + ui, family = binomial,  
    data = birthwt)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7308	-0.7841	-0.5144	0.9539	2.1980

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.125326	0.967561	-0.130	0.89694
lwt	-0.015918	0.006954	-2.289	0.02207 *
race2	1.300856	0.528484	2.461	0.01384 *
race3	0.854414	0.440907	1.938	0.05264 .
smoke1	0.866582	0.404469	2.143	0.03215 *
ptdTRUE	1.128857	0.450388	2.506	0.01220 *
ht1	1.866895	0.707373	2.639	0.00831 **
ui1	0.750649	0.458815	1.636	0.10183

### Final model

NB : the standard errors and p values are not correct because they still don't take into account model selection uncertainty.

It is generally NOT recommended to make inference on the best model without standard error adjustments but this is a common practice

# Classical approach to model selection

Model selection with stepwise AIC procedure :  
Better than multiple tests and very quick method

Problems :

- better to use AICc (possible with option k= ??)
- you not always find the best model  
(local optimum, different methods : backward, forward,...)
- model selection uncertainty

	model	k	n	AIC	AIC.delta
	lwt+ race+ smoke+ ptd+ ht+ ui	8	189	213.852	0.000
	lwt+ race+ smoke+ ptd+ ht	7	189	214.482	0.631
	age+ lwt+ race+ smoke+ ptd+ ht+ ui	9	189	214.834	0.982
	age+ lwt+ race+ smoke+ ptd+ ht	8	189	215.151	1.300
	lwt+ race+ smoke+ ptd+ ht+ ui+ ftv	9	189	215.838	1.986
	lwt+ race+ smoke+ ptd+ ht+ ftv	8	189	216.482	2.631
	lwt+ race+ ptd+ ht+ ui	7	189	216.567	2.715
	age+ lwt+ race+ smoke+ ptd+ ht+ ui+ ftv	10	189	216.750	2.899
	age+ lwt+ race+ smoke+ ptd+ ht+ ftv	9	189	217.116	3.264
	age+ lwt+ ptd+ ht+ ui	6	189	217.153	3.302

stepAIC found the best model in this case but other models have very close AIC values.

If you resample the data (eg with bootstrap) another model could have the lowest AIC value

**AIC.delta = AIC - AICmin**

Rule of thumb often used : models with AIC.delta  $\leq 2$   
are equally supported by the data



# Model selection uncertainty & likelihood of the model

It is easy to estimate the model selection uncertainty from the AICc value :

$$w_i = \exp(-0.5 * \text{AICc.delta}_i) / \sum(\exp(-0.5 * \text{AICc.delta}_i))$$

$$w_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_{r=1}^R \exp(-\frac{1}{2}\Delta_i)}$$

$W_i$  = model weight = Aikaike weight

= The probability for model i to be estimated as the best K-L model if we resample data (similar results obtained by bootstrapping)

= Likelihood (model<sub>i</sub> | data, set of models)

= Weight of evidence in favor of model i / hypothesis i  
ie : How much data support this model relative to other models

# Model selection uncertainty & likelihood of the model

## Aikaike weights , AICc.w :

```
> source("/home/gilles/stats/model.select_0.4.1.R")
> res <- model.select(mod)
> res$AICtab[1:20,]
```

	model	k	n	AICc	AICc.delta	AICc.w	sum.w
127	lwt+ race+ smoke+ ptd+ ht+ ui	8	189	214.652	0.000	0.104	0.104
63	lwt+ race+ smoke+ ptd+ ht	7	189	215.101	0.449	0.083	0.187
128	age+ lwt+ race+ smoke+ ptd+ ht+ ui	9	189	215.839	1.188	0.057	0.245
64	age+ lwt+ race+ smoke+ ptd+ ht	8	189	215.951	1.300	0.054	0.299
255	lwt+ race+ smoke+ ptd+ ht+ ui+ ftv	9	189	216.843	2.192	0.035	0.334
119	lwt+ race+ ptd+ ht+ ui	7	189	217.185	2.534	0.029	0.363
191	lwt+ race+ smoke+ ptd+ ht+ ftv	8	189	217.282	2.631	0.028	0.391
116	age+ lwt+ ptd+ ht+ ui	6	189	217.615	2.963	0.024	0.415
52	age+ lwt+ ptd+ ht	5	189	217.761	3.110	0.022	0.437
115	lwt+ ptd+ ht+ ui	5	189	217.840	3.189	0.021	0.458
60	age+ lwt+ smoke+ ptd+ ht	6	189	217.854	3.203	0.021	0.479
55	lwt+ race+ ptd+ ht	6	189	217.859	3.207	0.021	0.500
124	age+ lwt+ smoke+ ptd+ ht+ ui	7	189	217.859	3.208	0.021	0.521
123	lwt+ smoke+ ptd+ ht+ ui	6	189	217.927	3.276	0.020	0.541
256	age+ lwt+ race+ smoke+ ptd+ ht+ ui+ ftv	10	189	217.986	3.334	0.020	0.560
120	age+ lwt+ race+ ptd+ ht+ ui	8	189	218.047	3.395	0.019	0.579
192	age+ lwt+ race+ smoke+ ptd+ ht+ ftv	9	189	218.122	3.470	0.018	0.598
59	lwt+ smoke+ ptd+ ht	5	189	218.283	3.632	0.017	0.615
51	lwt+ ptd+ ht	4	189	218.341	3.689	0.016	0.631
56	age+ lwt+ race+ ptd+ ht	7	189	218.391	3.739	0.016	0.647

+ 236 other models ...

k = number of parameters in the model

n = number of observations : **n MUST be equal in all models !**

model.select  
computes  
automatically all  
possible  
sub-models from a  
full model  
→ no more  
problems of local  
optimum

sum.w :  
cumulative sum of  
AICc.w = 1

# Model selection uncertainty & likelihood of the model

## Use of Aikake weights :

	model	k	n	AICc	AICc.delta	AICc.w	sum.w
127	lwt+ race+ smoke+ ptd+ ht+ ui	8	189	214.652	0.000	0.104	0.104
63	lwt+ race+ smoke+ ptd+ ht	7	189	215.101	0.449	0.083	0.187
128	age+ lwt+ race+ smoke+ ptd+ ht+ ui	9	189	215.839	1.188	0.057	0.245
64	age+ lwt+ race+ smoke+ ptd+ ht	8	189	215.951	1.300	0.054	0.299
255	lwt+ race+ smoke+ ptd+ ht+ ui+ ftv	9	189	216.843	2.192	0.035	0.334
(...)							

Model 1 has a probability of 0.104 to be selected as the best model  
ie : if you collect 1000 new datasets we estimate that model 1 will have the lowest AICc of  
the set of models in 104 cases

## Evidence ratios :

Model 1 is ~ 3 times more supported by the data than model 5  
( $w_1/w_5 = 0.104 / 0.035 = 2.97$ )  
cfr rule of thumb delta AIC  $\leq 2$

## Confidence set of models

The best K-L model (for this dataset and this set of models)  
has a probability of 0.334 to be within the 5 first models ( $sum.w$ )<sub>35</sub>

# Model selection uncertainty & likelihood of the model

NB : the more models you have in the set of models, the higher the uncertainty → it is generally better to restrict the number of models as much as possible

Two main approaches with these AIC methods :

1) select a **restricted number of carefully chosen models** corresponding to the biological hypotheses you want to compare (with AICc weights)

2) Start from one full model with all potentially important explanatory variables and compute "**all possible models**" from the combinations of these explanatory variables.

→

generally higher uncertainty but this approach is more adapted for the following interesting methods

# Estimating the relative importance of predictor variables

$w_+(j) = \text{Sum } (w_i)$  of the models in which variable  $j$  is present

```
> source("/home/gilles/stats/R/WorkingDirectory/rprojects/model.select_0.4.R")
> res <- model.select(mod)

> res$AICtab[1:20, c("model", "k", "n", "AICc", "AICc.delta", "AICc.w", "sum.w")]

      model k  n  AICc AICc.delta AICc.w sum.w
127      lwt+ race+ smoke+ ptd+ ht+ ui  8 189 214.652  0.000  0.104 0.104
63      lwt+ race+ smoke+ ptd+ ht  7 189 215.101  0.449  0.083 0.187
128     age+ lwt+ race+ smoke+ ptd+ ht+ ui  9 189 215.839  1.188  0.057 0.245
64      age+ lwt+ race+ smoke+ ptd+ ht  8 189 215.951  1.300  0.054 0.299
255     lwt+ race+ smoke+ ptd+ ht+ ui+ ftv  9 189 216.843  2.192  0.035 0.334
119      lwt+ race+ ptd+ ht+ ui  7 189 217.185  2.534  0.029 0.363
191     lwt+ race+ smoke+ ptd+ ht+ ftv  8 189 217.282  2.631  0.028 0.391
116      age+ lwt+ ptd+ ht+ ui  6 189 217.615  2.963  0.024 0.415
+ 250 other models ...
```

```
> res$var.weights

      (Intercept)  freq  w
      ptd          0.5 0.950
      ht           0.5 0.894
      lwt          0.5 0.842
      smoke        0.5 0.711
      race         0.5 0.711
      ui           0.5 0.551
      age          0.5 0.445
      ftv          0.5 0.257
```

freq = proportion of models in which the parameter is present  
Ideally the initial frequency should be the same for all parameters.

This can be achieved with an "all possible models" approach

# Inference from multiple models

We have now nice solutions for several problems :

If we estimate all possible models

We don't have a risk of local optimum or different final model depending on the method used (backward, forward,...)

But this approach has also drawbacks, ie : increased uncertainty in model selection, impossible to do with very complex models (many x)...

We have interpretable values providing the degree of support by the data of the different models/hypotheses (AICc.w) and of the different explanatory variable (w+)

No problem with multiple testing as AIC is a direct measure of model quality (in K-L distance term)

# Inference from multiple models

But we still have often several models that are almost equivalently supported by the data (delta AIC close to 0).

**Which model should we use** (eg for prediction) ?

A common practice is to **consider the models with delta AICc < 2** and use the most parsimonious one (ie with the smallest number of parameters).

But by doing this you drop information that could make sense in your system and you still have standard errors that do not take into account the model selection process (they are conditional to the model)

Another solution is to use the information from all models at once thanks to **model averaging** of the coefficients and unconditional standard error estimates

# Inference from multiple models

## Model averaged coefficients :

= mean of the coefficient value in all models weighed by the AICc weight of each model.

$$\tilde{\beta}_j = \sum_{i=1}^R w_i I_j(g_i) \hat{\beta}_{j,i}$$

## Unconditional Standard Errors :

= standard errors that take into account the model selection uncertainty ie that are not conditional to the model, but to the set of models considered

$$\hat{var}(\tilde{\theta}_i) = \left( \sum_{i=1}^R w_i \sqrt{\hat{var}(\hat{\theta}_i | g_i) + (\hat{\theta}_i - \tilde{\theta}_i)^2} \right)^2$$

	model	k	n	AICc	AICc.delta	AICc.w	sum.w
127	lwt+ race+ smoke+ ptd+ ht+ ui	8	189	214.652	0.000	0.104	0.104
63	lwt+ race+ smoke+ ptd+ ht	7	189	215.101	0.449	0.083	0.187
128	age+ lwt+ race+ smoke+ ptd+ ht+ ui	9	189	215.839	1.188	0.057	0.245
64	age+ lwt+ race+ smoke+ ptd+ ht	8	189	215.951	1.300	0.054	0.299
255	lwt+ race+ smoke+ ptd+ ht+ ui+ ftv	9	189	216.843	2.192	0.035	0.334
119	lwt+ race+ ptd+ ht+ ui	7	189	217.185	2.534	0.029	0.363
191	lwt+ race+ smoke+ ptd+ ht+ ftv	8	189	217.282	2.631	0.028	0.391
116	age+ lwt+ ptd+ ht+ ui	6	189	217.615	2.963	0.024	0.415
+ 250 other models ...							

Unconditional variance covariance matrix not (yet) implemented



# Inference from multiple models

## Model averaged coefficients & Unconditional Standard Errors

```

                                model k   n   AICc AICc.delta AICc.w sum.w
127          lwt+ race+ smoke+ ptd+ ht+ ui  8 189 214.652      0.000  0.104 0.104
63              lwt+ race+ smoke+ ptd+ ht  7 189 215.101      0.449  0.083 0.187
128    age+ lwt+ race+ smoke+ ptd+ ht+ ui  9 189 215.839      1.188  0.057 0.245
64          age+ lwt+ race+ smoke+ ptd+ ht  8 189 215.951      1.300  0.054 0.299
255    lwt+ race+ smoke+ ptd+ ht+ ui+ ftv  9 189 216.843      2.192  0.035 0.334
119              lwt+ race+ ptd+ ht+ ui  7 189 217.185      2.534  0.029 0.363
191    lwt+ race+ smoke+ ptd+ ht+ ftv  8 189 217.282      2.631  0.028 0.391
116              age+ lwt+ ptd+ ht+ ui  6 189 217.615      2.963  0.024 0.415
+ 250 other models ...

```

```

> res$mod.av[, c("freq", "w", "av.coef", "av.se")]
      freq      w av.coef av.se
(Intercept) 1.0 1.000   0.386 1.297
ptdTRUE      0.5 0.950   1.235 0.443
ht1          0.5 0.894   1.587 0.669
lwt          0.5 0.842  -0.013 0.006
race2       0.5 0.711   0.834 0.450
race3       0.5 0.711   0.559 0.365
smoke1      0.5 0.711   0.576 0.340
ui1         0.5 0.551   0.409 0.313
age         0.5 0.445  -0.021 0.021
ftv         0.5 0.257  -0.002 0.045

```

You can compute an approximate unconditional 95% confidence interval with  $\text{av.coef} \pm 1.96 * \text{av.se}$

NB : these se are generally much higher than the conditional se  
 The higher the model selection uncertainty, the higher av.se.  
 → you should limit the number of explanatory variables

# Inference from multiple models

There are in fact 2 slightly different way to calculate the averaged coefficients

"Normal" model averaging

$$\hat{\beta}_j = \frac{\sum_{i=1}^R w_i I_j(g_i) \hat{\beta}_{j,i}}{\sum_{i=1}^R w_i I_j(g_i)}$$

$$\hat{var}(\hat{\theta}_i) = \left( \sum_{i=1}^R \frac{w_i}{\sum_i w_i I_j(g_i)} \sqrt{\hat{var}(\hat{\theta}_i | g_i) + (\hat{\theta}_i - \hat{\theta})^2} \right)^2$$

Use only models in which beta\_j is present  
Rescale the w\_i so that their sum is = 1

"Shrinkage" model averaging

$$\tilde{\beta}_j = \sum_{i=1}^R w_i I_j(g_i) \hat{\beta}_{j,i}$$

$$\hat{var}(\tilde{\theta}_i) = \left( \sum_{i=1}^R w_i \sqrt{\hat{var}(\hat{\theta}_i | g_i) + (\hat{\theta}_i - \tilde{\theta})^2} \right)^2$$

Use all models  
When beta\_j is not present its value is 0

→ estimates shrunk towards 0 for  
parameters present mainly in "bad models"  
ie parameters with low w+

# Inference from multiple models

```
> res$mod.av
      freq      w av.coef av.se av.coef2 av.se2
(Intercept) 1.0 1.000  0.386 1.297   0.386 1.297
ptdTRUE      0.5 0.950  1.235 0.443   1.299 0.462
ht1          0.5 0.894  1.587 0.669   1.775 0.724
lwt          0.5 0.842  -0.013 0.006  -0.016 0.007
race2        0.5 0.711   0.834 0.450   1.174 0.538
race3        0.5 0.711   0.559 0.365   0.787 0.464
smoke1       0.5 0.711   0.576 0.340   0.809 0.421
ui1          0.5 0.551   0.409 0.313   0.743 0.460
age          0.5 0.445  -0.021 0.021  -0.048 0.038
ftv          0.5 0.257  -0.002 0.045  -0.010 0.176
```

shrinkage model  
averaging

normal model  
averaging

the difference is  
higher for poorly  
supported  
variables  
av.coef is more  
shrunk toward 0

## "Normal" model averaging :

- not consistent with averaging of the predictions
- parameters from the same model will be weighed differently  
(don't seem to be a good thing, ie for interactions)

- shrinkage seems to be an interesting property :

when no shrinkage : a parameter that is only in "bad models" can have a high size effect

→ we prefer shrinkage model averaging (av.coef instead of av.coef2<sup>43</sup>)

# Problems

- 1) Qaic
- 2) Application in Mixed models
- 3) Interpretation with interactions or in case of non balanced set of models

## Problems : QAIC

Coefficient of overdispersion estimated from the full model and used in all other models

Do not use the coefficient of each model to calculate the QAIC !

→ strange ?

→ you need a full model (not always desirable)

What to do with standard errors for model averaging?

→ MuMin package : se taken as it

→ model.select : se errors multiplied by sqrt of overdispersion coefficient before model averaging

# Problems : Mixed models

AIC not adapted to compare models with different random structures  
(ie because parameters "tested" at the margin)

But frequently used nevertheless in practice (because best available option ?)

→ Compare models with different fixed effects but the same random structure

Ok if no interest in random effects that are just "nuisance effects"

**!!! use ML, not REML to compare models with different fixed structures !!!**

What are the number of parameters  
(K used to calculate AIC)?

If interest only in the fixed effect part :

ok to consider each random effect (hyper parameter) as 1 parameter

It is possible to avoid the use of QAIC (at least with R lmer package)

Modelise the overdispersion in the random effects (observation level random effect)

# Problems : when there is no balance

If the frequency of the explanatory variables in the initial set of model is not the same

(typically when you include interaction and respect marginality)

→ you can't compare the weights of the different parameters ?

Empirical rule of thumb ? :

explanatory variables with weight  $>$  initial frequency are important, supported by the data

If weight  $<$  frequency : not supported

If weight  $\sim$  frequency : small, borderline , not very well supported effects

```
> mod <- glm(low ~ (age + lwt + smoke + ht + smoke:lwt + smoke:ht),
+ family=binomial, data=birthwt)
>
> res <- model.select(mod)

> res$var.weights[, -2]
      freq      w
(Intercept) 1.000 1.000
lwt          0.615 0.964
ht           0.615 0.932
smoke        0.692 0.858
age          0.500 0.389
lwt:smoke    0.231 0.297
smoke:ht     0.231 0.208
```

# Problems : when there is no balance

Other option :

Null hypothesis testing based on randomization of the response Y

P value = Probability to obtain a W as high or higher by pure hasard, ie if the variable has no explanatory power (ie when the Y are randomized)  
(NB : method discussed by Burnham & Anderson)

Implemented in the `model.select` function but much more computer intensive.

Only for lm and glm.

For mixed models, you must make the randomization yourself and pay attention to keep the random structure (e.g. randomize within groups).

```
> mod <- glm(low ~ (age + lwt + smoke + ht + smoke:lwt + smoke:ht),  
+ family=binomial, data=birthwt)  
>  
>  
res <- model.select(mod, nsimul=100)
```

```
> res$var.weights[,-2]  
      freq      p      w  
(Intercept) 1.000 1.00 1.000  
lwt          0.615 0.00 0.964  
ht           0.615 0.01 0.932  
smoke        0.692 0.04 0.858  
age          0.500 0.32 0.389  
lwt:smoke    0.231 0.07 0.297  
smoke:ht     0.231 0.08 0.208
```

p values by  
randomization







# Practical use of `model.select`

An example with Birthwt dataset from MASS package

NB : we will focus here on the use of `model.select` (options etc...)  
The models used are probably crazy here.

NB2 : `model.select_0.3.R` has been used and tested during several years without problems. It worked well with `lm`, `glm`, `nlme` and `lme4 0.999xx`  
`model.select_0.4.R` is (was : 2013) a very recent version transformed to be used with `lme4 1.0.x`  
The output structure has slightly been modified  
It has not been tested as extensively → be prudent with the results (as always)

Don't hesitate to use the CRAN packages `MuMin` and `AICcmodavg` in stead  
Comparison `model.select` - `MuMin` here :  
<http://forums.cirad.fr/logiciel-R/viewtopic.php?t=3517>

```
library(MASS)
data(birthwt)

# tranform some variables into factors
birthwt[,c("race", "smoke", "ht", "ui")] <-
  lapply(birthwt[,c("race", "smoke", "ht", "ui")], factor)
birthwt$ptd <- factor(birthwt$ptd > 0)

# compute a binomial model
mod <- glm(low ~ (age + lwt + race + smoke), family=binomial,
data=birthwt)
```

# Practical use of model.select

First kind of use : all possible models from a full model

```
> source("/home/gilles/stats/R/model.select_0.4.R")
> res <- model.select(mod)
> res
$AICtab
      model id k   n   loglik   AICc AICc.delta AICc.w sum.w
15    lwt+ race+ smoke 15 5 189 -107.5073 225.343      0.000 0.474 0.474
16 age+ lwt+ race+ smoke 16 6 189 -107.2886 227.039      1.696 0.203 0.677
13                race+ smoke 13 4 189 -109.9874 228.192      2.850 0.114 0.791
14    age+ race+ smoke 14 5 189 -109.4311 229.190      3.848 0.069 0.860
(...)
1                1 1 189 -117.3360 236.693     11.351 0.002 1.000
```

\$var.weights

	freq	w
(Intercept)	1.0	1.000
smoke	0.5	0.938
race	0.5	0.897
lwt	0.5	0.793
age	0.5	0.331

AICc table

\$mod.av

	freq	w	av.coef	av.se	av.coef2	av.se2
(Intercept)	1.0	1.000	-0.156	1.137	-0.156	1.137
smoke1	0.5	0.938	0.973	0.370	1.037	0.388
race3	0.5	0.897	0.869	0.391	0.969	0.423
race2	0.5	0.897	1.093	0.477	1.218	0.517
lwt	0.5	0.793	-0.010	0.005	-0.013	0.006
age	0.5	0.331	-0.009	0.013	-0.028	0.035

variable weights w+

shrinkage model averaging

normal model averaging

unconditional standard errors

# Practical use of `model.select`

Second kind of use : a list of carefully chosen models

```
> modlist <- list(
+   mod1 = glm(low ~ age ,family=binomial, data=birthwt),
+   mod2 = glm(low ~ age + lwt ,family=binomial, data=birthwt),
+   mod3 = glm(low ~ age + lwt + age:lwt,family=binomial, data=birthwt),
+   mod4 = glm(low ~ I(age/sqrt(lwt)) ,family=binomial, data=birthwt),
+   mod5 = glm(low ~ lwt, family=binomial, data=birthwt),
+   mod6 = glm(low ~ 1, family=binomial, data=birthwt)
+ )
```

```
> model.select(model = NULL, models.list = modlist)
```

\$AICtab

	model	id	k	n	loglik	AICc	AICc.delta	AICc.w	sum.w
mod5	lwt	5	2	189	-114.3453	232.755	0.000	0.408	0.408
mod2	age+ lwt	2	3	189	-113.5617	233.253	0.498	0.318	0.727
mod3	age+ lwt+ age:lwt	3	4	189	-113.5616	235.341	2.585	0.112	0.839
mod1	age	1	2	189	-115.9560	235.976	3.221	0.082	0.921
mod6		6	1	189	-117.3360	236.693	3.938	0.057	0.978
mod4	I(age/sqrt(lwt))	4	2	189	-117.2468	238.558	5.803	0.022	1.000

# Practical use of `model.select`

option `srt` : choose an other Information Criterion than the default AICc  
Possible values = AICc, AIC, QAICc, QAIC

```
> model.select(mod, srt = "QAICc")
$AICtab
      model id k   n   loglik   QAICc QAICc.delta QAICc.w sum.w
15    lwt+ race+ smoke 15 5 189 -107.5073 195.835      0.000  0.384 0.384
16  age+ lwt+ race+ smoke 16 6 189 -107.2886 197.619      1.784  0.157 0.542
13                race+ smoke 13 4 189 -109.9874 197.931      2.096  0.135 0.676
(...)
1                1 1 189 -117.3360 204.202      8.368  0.006 0.996
6                age+ race  6 4 189 -114.0638 204.884      9.049  0.004 1.000
(...)

$mod.av
      freq      w av.coef av.se av.coef2 av.se2
(Intercept)  1.0 1.000  -0.144 1.242  -0.144  1.242
smoke1       0.5 0.884   0.896 0.389   1.014  0.423
race3        0.5 0.809   0.778 0.403   0.962  0.461
race2        0.5 0.809   0.970 0.487   1.199  0.559
lwt          0.5 0.734  -0.010 0.006  -0.013  0.007
age          0.5 0.332  -0.010 0.014  -0.031  0.037
```

# Practical use of `model.select`

option `keep` : force one explanatory variable to be in all models.  
Useful if you want to always control for a covariable

```
> model.select(mod, keep=c("race"))
```

```
$AICtab
```

	model	id	k	n	loglik	AICc	AICc.delta	AICc.w	sum.w
7	lwt+ race+ smoke	7	5	189	-107.5073	225.343	0.000	0.528	0.528
8	age+ lwt+ race+ smoke	8	6	189	-107.2886	227.039	1.696	0.226	0.755
5	race+ smoke	5	4	189	-109.9874	228.192	2.850	0.127	0.882
6	age+ race+ smoke	6	5	189	-109.4311	229.190	3.848	0.077	0.959
3	lwt+ race	3	4	189	-111.6295	231.476	6.134	0.025	0.983
4	age+ lwt+ race	4	5	189	-111.3303	232.989	7.646	0.012	0.995
1	race	1	3	189	-114.8308	235.791	10.449	0.003	0.998
2	age+ race	2	4	189	-114.0638	236.345	11.002	0.002	1.000

```
$var.weights
```

	freq	w
(Intercept)	1.0	1.000
race	1.0	1.000
smoke	0.5	0.959
lwt	0.5	0.791
age	0.5	0.317

```
(...)
```

# Practical use of `model.select`

option `nvmax` : maximum number of variables that you want in the models. (avoid to compute overly complex models)

```
> model.select(mod, nvmax = 2)
```

```
$AICtab
```

	model	id	k	n	loglik	AICc	AICc.delta	AICc.w	sum.w
11	race+ smoke	11	4	189	-109.9874	228.192	0.000	0.526	0.526
10	lwt+ smoke	10	3	189	-112.1703	230.470	2.278	0.168	0.694
7	lwt+ race	7	4	189	-111.6295	231.476	3.284	0.102	0.796
3	lwt	3	2	189	-114.3453	232.755	4.563	0.054	0.850
4	age+ lwt	4	3	189	-113.5617	233.253	5.061	0.042	0.892
9	age+ smoke	9	3	189	-113.6381	233.406	5.214	0.039	0.930
8	smoke	8	2	189	-114.9023	233.869	5.677	0.031	0.961
5	race	5	3	189	-114.8308	235.791	7.599	0.012	0.973
2	age	2	2	189	-115.9560	235.976	7.784	0.011	0.984
6	age+ race	6	4	189	-114.0638	236.345	8.153	0.009	0.993
1		1	1	189	-117.3360	236.693	8.501	0.007	1.000

```
$var.weights
```

	freq	w
(Intercept)	1.000	1.000
smoke	0.364	0.764
race	0.364	0.648
lwt	0.364	0.366
age	0.364	0.100

# Practical use of `model.select`

option `null.model` : logical indicating if you want to include a model with just the intercept in the set of models (default = TRUE)

```
> model.select(mod, null.model=FALSE)
```

```
$AICtab
```

				model id	k	n	loglik	AICc	AICc.delta	AICc.w	sum.w
14	lwt+	race+	smoke	14 5	189	-107.5073	225.343	0.000	0.475	0.475	
15	age+	lwt+	race+	smoke	15 6	189	-107.2886	227.039	1.696	0.203	0.678
12		race+	smoke	12 4	189	-109.9874	228.192	2.850	0.114	0.792	
13	age+	race+	smoke	13 5	189	-109.4311	229.190	3.848	0.069	0.862	
10		lwt+	smoke	10 3	189	-112.1703	230.470	5.128	0.037	0.898	
11	age+	lwt+	smoke	11 4	189	-111.4397	231.097	5.754	0.027	0.925	
6		lwt+	race	6 4	189	-111.6295	231.476	6.134	0.022	0.947	
2		lwt		2 2	189	-114.3453	232.755	7.413	0.012	0.959	
7	age+	lwt+	race	7 5	189	-111.3303	232.989	7.646	0.010	0.969	
3		age+	lwt	3 3	189	-113.5617	233.253	7.911	0.009	0.978	
9		age+	smoke	9 3	189	-113.6381	233.406	8.063	0.008	0.986	
8		smoke		8 2	189	-114.9023	233.869	8.527	0.007	0.993	
4		race		4 3	189	-114.8308	235.791	10.449	0.003	0.996	
1		age		1 2	189	-115.9560	235.976	10.634	0.002	0.998	
5		age+	race	5 4	189	-114.0638	236.345	11.002	0.002	1.000	

```
$var.weights
```

	freq	w
(Intercept)	1.000	1.000
smoke	0.533	0.940
race	0.533	0.899
lwt	0.533	0.795
age	0.533	0.332



# Practical use of `model.select`

option `nsimul` : number of permutations in the permutation tests of the variables weights (default = 0)

```
> mod <- glm(low ~ (age + lwt + smoke)^2, family=binomial, data=birthwt)
> model.select(mod, nsimul=100)
$AICtab
```

	model	id	k	n	loglik	AICc	AICc.delta	AICc.w	sum.w
7	lwt+ smoke	7	3	189	-112.1703	230.470	0.000	0.162	0.162
14	lwt+ smoke+ lwt:smoke	14	4	189	-111.1859	230.589	0.119	0.153	0.315
8	age+ lwt+ smoke	8	4	189	-111.4397	231.097	0.626	0.119	0.434
15	age+ lwt+ smoke+ lwt:smoke	15	5	189	-110.5596	231.447	0.977	0.100	0.534
12	age+ lwt+ smoke+ age:smoke	12	5	189	-110.8160	231.960	1.489	0.077	0.611
(...)									
2	age	2	2	189	-115.9560	235.976	5.506	0.010	0.993
1		1	1	189	-117.3360	236.693	6.223	0.007	1.000

```
$var.weights
```

	freq	nullw	p	w
(Intercept)	1.000	1.000	1.00	1.000
lwt	0.722	0.476	0.05	0.890
smoke	0.722	0.432	0.04	0.876
age	0.722	0.429	0.18	0.596
lwt:smoke	0.278	0.076	0.05	0.358
age:smoke	0.278	0.075	0.14	0.199
age:lwt	0.278	0.083	0.25	0.135