Lecture 8: Model/variable selection

BIO144 Data Analysis in Biology

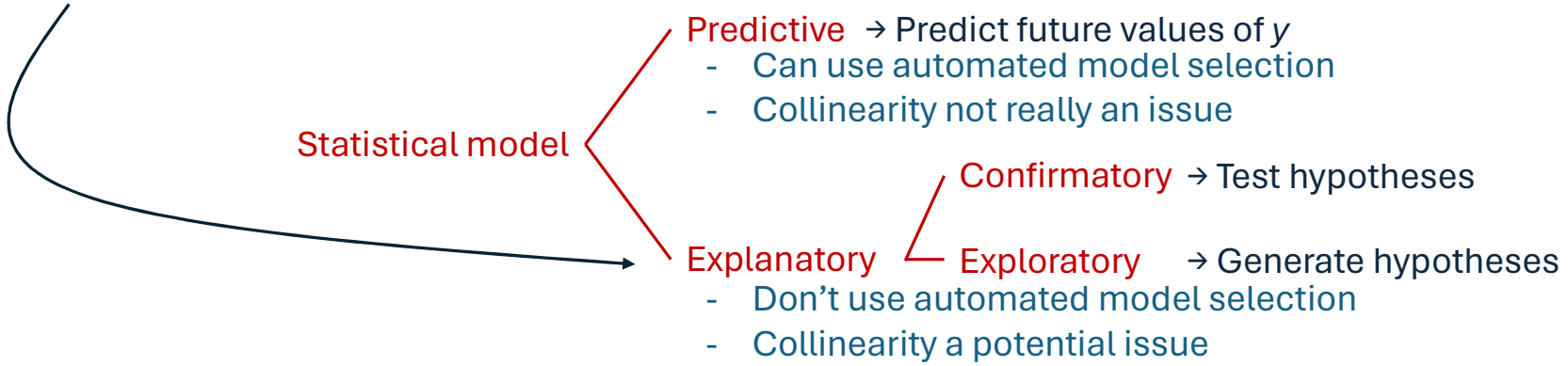
Owen Petchey, Stephanie Muff, Erik Willems

University of Zurich

22 April 2024

Overview

- Predictive vs. explanatory models
- Selection criteria: AIC, AIC_c, and BIC
- Automatic model selection and its caveats
- Model selection bias
- Collinearity among explanatory variables
- Occam's razor



Course material covered today

Today's lecture material is partially based on the following literature:

- "Lineare regression" chapters 5.1-5.4
- Chapter 27.1 and 27.2 by Clayton and Hills "Choice and Interpretation of Models" (pdf provided)

Optional reading:

- Paper by Freedman 1983: "A Note on Screening Regression Equations" (Sections 1 and 2 are sufficient to get the point)

Developing a model

So far, our regression models "fell from heaven": The model family and the terms in the model were almost always given.

However, it is often not immediately obvious which terms are relevant to include in a model.

Importantly, the approach to formulate a model **heavily depends on the aim** for which the model is built.

The following distinction is important:

- The aim is to predict future values of **y** from known regressors. Variables in the model are covariates.
- The aim is to explain **y** using known regressors. Ultimately, the aim is to find causal relationships. Variables in the model are explanatory variables.

→ Even among statisticians there is no real consensus about how, if, or when to select a model:

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016, **7**, 679–692 doi: 10.1111/2041-210X.12541

SPECIAL FEATURE: 5TH ANNIVERSARY OF *METHODS IN ECOLOGY AND EVOLUTION*

The relative performance of AIC, AIC_C and BIC in the presence of unobserved heterogeneity

Mark J. Brewer[1,*], Adam Butler[2] and Susan L. Cooksley[3]

[1]*Biomathematics and Statistics Scotland, Craigiebuckler, Aberdeen, AB15 8QH,UK;* [2]*Biomathematics and Statistics Scotland, JCMB, The King's Buildings, Edinburgh, EH9 3JZ,UK; and* [3]*The James Hutton Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK*

Summary

1. Model selection is difficult. Even in the apparently straightforward case of choosing between standard linear regression models, there does not yet appear to be consensus in the statistical ecology literature as to the right approach.

Note: The first sentence of a paper in *Methods in Ecology and Evolution* from 2016 is: "Model selection is difficult."

Why is finding a model so hard?

Remember from week 1:

> Ein Modell ist eine Annäherung an die Realität. Das Ziel der Statistik und Datenanalyse ist es immer, dank Vereinfachungen der wahren Welt gewisse Zusammenhänge zu erkennen.

Box (1979): "All models are wrong, but some are useful."

→ There is often not a "right" or a "wrong" model – but there are more and less useful ones.

→ Finding a model with good properties is sometimes an art...

Predictive and explanatory models

Before we continue to discuss model/variable selection, we need to be clear about the scope of the model:

- **Predictive models**: These are models that aim to predict the outcome of future subjects.
  Example: In the bodyfat example the aim is to predict people's bodyfat from factors that are easy to measure (age, BMI, weight,..).
- **Explanatory models**: These are models that aim at understanding the (causal) relationship between explanatory variables and the response.
  Example: The mercury study aims to understand if Hg-concentrations in the soil (explanatory) influence the Hg-concentrations in humans (response).

→ The model selection strategy depends on this distinction.

Prediction vs explanation

When the aim is **prediction**, the best model is the one that best predicts the value of the outcome for a future subject. This is a well defined task and "objective" variable selection strategies to find the model which is best in this sense are potentially useful.

However, when used for **explanation** the best model will depend on the scientific question being asked, **and automatic variable selection strategies have no place**.

(Clayton and Hills, 1993, chapters 27.1 and 27.2)

Use your "domain knowledge" (medical, biological, ...) instead!

A predictive model: The bodyfat example

The bodyfat study is a typical example of a **predictive model**.

There are 12 potential predictors for the response variable. Let's fit the full model (without interactions):

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-115.957	57.197	-2.027	0.044
age	0.020	0.031	0.643	0.521
gewicht	-0.763	0.353	-2.164	0.032
hoehe	0.584	0.318	1.836	0.068
bmi	2.481	1.127	2.202	0.029
neck	-0.598	0.224	-2.670	0.008
chest	-0.144	0.112	-1.286	0.200
abdomen	0.923	0.094	9.876	0.000
hip	-0.309	0.154	-2.007	0.046
thigh	0.248	0.153	1.625	0.106
knee	0.073	0.258	0.285	0.776
ankle	-0.490	0.343	-1.429	0.154
biceps	0.165	0.167	0.987	0.324

Don't interpret!

Caricature of approach:

- "throw them all in"
- Let the data speak for themselves

→ the (eventual) model is a blackbox: *prediction without understanding*

Model selection for predictive models

- Remember: R^2 is not suitable for model selection, because it *always* increases (improves) when a new variable is included.
- Ideally, the predictive ability of a model is tested by a cross-validation (CV) approach. ▸ Find a description of the CV idea here.
- CV can be a bit cumbersome, and sometimes would require additional coding.
- Approximations to CV: So-called information-criteria like AIC, AIC_c, BIC...
- The idea is that the "best" model is the one with the smallest value of the information criterion (where the criterion is selected in advance).

Information-criteria

Information-criteria for model selection were made popular by Burnham & Anderson (2002)

The idea is to find a balance between

Good model fit $\leftrightarrow$ **Low model complexity**

$\rightarrow$ Reward models with a better fit to the data.

$\rightarrow$ Penalize models with more parameters.

AIC

The most prominent criterion is the AIC (Akaike Information Criterion), which measures the relative quality of a model.

The AIC of a model with likelihood L and p parameters is given as:

$$AIC = -2\log(L) + 2p$$

Important: The lower the AIC, the better the model!

The AIC is a trade-off between:

- a high likelihood L (good model fit)
- few parameters p (low model complexity)

AIC_c: The AIC for low sample sizes

When the number of data points n is small with respect to the number of parameters p in a model, the use of a corrected AIC, the AIC_c is recommended.

The **corrected AIC** of a model with n data points, likelihood L and p parameters is given as:

$$AIC_c = -2\log(L) + 2p \cdot \frac{n}{n-p-1}$$

Burnham and Anderson **recommend to use AIC_c** in general, especially **when** $n/p < 40$.

Penalty term is higher than in the normal AIC, especially for smaller sample sizes

In the bodyfat example, we have 243 data points and 13 parameters (including the intercept β_0), thus $n/p = 143/13 \approx 19 < 40 \Rightarrow$ use AIC_c for model selection!

BIC, the brother/sister of AIC

Other information criteria were suggested as well. Another prominent example is the BIC (Bayesian Information Criterion), which is similar in spirit to the AIC.

The BIC of a model for n data points with likelihood L and p parameters is given as:

$$BIC = -2\log(L) + p \cdot \ln(n)$$

Again: The lower the BIC, the better the model!

The only difference to AIC is the penalty for model complexity.

Model selection with AIC/AICc/BIC

Given m potential variables to be included in a model.

- In principle it is possible to minimize the AIC/AICc/BIC over all 2^m possible models. Simply fit all models and take the "best" one (lowest AIC).
- This is cumbersome to do "by hand". Useful to rely on implemented procedures in R, which search for the model with the lowest AIC/AICc/BIC.

- Backward selection: **Start with a large/full model.** In each step, **remove** the variable that leads to the largest improvement (smallest AIC/AICc/BIC). Do this until no further improvement is possible.
- Forward selection: **Start with a null model.** In each step, **add** the predictor that leads to the largest improvement (smallest AIC/AICc/BIC). Do this until no further improvement is possible.

"Best" predictive model for bodyfat

Given the predictive nature of the bodyfat model, we search for the model with minimal AICc, for instance using the `stepAIC()` function from the `MASS` package:

```
library(MASS)
library(AICcmodavg)
# Remember: r.bodyfat <- lm(bodyfat ~ ., d.bodyfat)
r.AIC <- stepAIC(r.bodyfat, direction= c("both"), trace= F, AICc= T)
AICc(r.bodyfat)
```

```
## [1] 1413.99
```

```
AICc(r.AIC)
```

```
## [1] 1408.469
```

$\rightarrow \Delta$AICc$=$ 5.52 in favour of `r.AIC`.

Information criteria express *relative* performance:

- Absolute value has no real interpretation
- Only compare models fitted to the same data (always check sample size!)

Note: Owen will also use `direction=c("forward")` and `direction=c("backward")` in the BC videos.

The model was reduced, with only 8 of the 12 initial predictor variables retained.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-112.867	55.770	-2.024	0.044
gewicht	-0.746	0.343	-2.176	0.031
hoehe	0.535	0.315	1.699	0.091
bmi	2.170	1.096	1.979	0.049
neck	-0.539	0.217	-2.479	0.014
abdomen	0.910	0.078	11.673	0.000
hip	-0.292	0.149	-1.966	0.050
thigh	0.301	0.131	2.297	0.023
ankle	-0.453	0.321	-1.414	0.159

Don't interpret!

Note 1: AICc minimization may lead to a model that retains variables with relatively large p-values (*e.g.*, ankle).

Note 2: We could continue and *e.g.* include interactions, transformations, *etc.*

Cautionary note about the "best" predictive model

It is tempting to look at the coefficients and try to interpret what you see, in the sense of "Increasing the weight by 1kg will cause a bodyfat reduction by -0.75 percentage points."

However, the coefficients of such an optimized "best" model should **not be interpreted**!

Really think of prediction optimised models as a blackbox: *prediction without understanding*

$\rightarrow$ Model selection may lead to biased parameter estimates: do not draw (biological, medical,..) conclusions from models that were optimized for prediction

See *e.g.*, Freedman1983, Copas 1983.

Your aim is explanation?

"Explanation" means you want to interpret the regression coefficients, 95% CIs, and p-values. It is then often assumed that some sort of causality ($x \rightarrow y$) exists.

In such a situation, you should formulate a confirmatory model:

- **Start with a clear hypothesis**
- **Select your explanatory variables according to** a priori knowledge.
- Ideally formulate **only one** or a few model(s) **before you start analysing your data.**

Confirmatory models have a long tradition in medicine. In fact, the main conclusions in a study are only allowed to be drawn from the main model (which needs to be specified even before data are collected):

It will rarely be necessary to include a large number of variables in the analysis, because only a few exposures are of genuine scientific interest in any one study, and there are usually very few variables of sufficient *a priori* importance for their potential confounding effect to be controlled for. Most scientists are aware of the dangers of analyses which search a long list of potentially relevant exposures. These are known as *data dredging* or *blind fishing* and carry a considerable danger of false positive findings. Such analyses are as likely to impede scientific progress as to advance it. There are similar dangers if a long list of potential confounders is searched, either with a view to explaining the observed relationship between disease and exposure or to enhancing it — findings will inevitably be biased. Confounders should be chosen *a priori* and not on the basis of statistical significance. In particular, variables which have been used in the design, such as matching variables, must be included in the analysis.

(chapters 27.1 and 27.2, Clayton Hills 1993)

Confirmatory vs. exploratory

Any **additional analyses** that you potentially do with your data are **exploratory**.

→ Two types of **explanatory models/analyses**:

- Confirmatory:
  - Clear hypothesis and **a priori** selection of regressors for y.
  - **No variable selection!**
  - Allowed to interpret the results and draw quantitative conclusions.
- Exploratory:
  - Build whatever model you want, but the results should only be used to generate new hypotheses, *a.k.a.* "speculations".
  - Clearly report the results as "exploratory".

Interpretation of exploratory models?

Results from exploratory models can be used to generate new hypotheses, not to draw causal conclusions, or to over-interpret effect-sizes.

In biological publications it is (unfortunately) still common practice to draw conclusions from exploratory models, optimized using model selection criteria (like AIC), as if the models were confirmatory!

$\rightarrow$ We illustrate why this is a problem with a (simulation) example on the next slides.

Model selection bias

Aim of the example:

To illustrate how model selection based on AIC can lead to biased parameters, and inflated effect sizes.

Procedure:

1. Randomly generate 100 data points for 50 explanatory variables $x^{(1)}, \ldots, x^{(50)}$, and a response variable y:

```
set.seed(123456)
data<- data.frame(matrix(rnorm(100 * 51), ncol= 51))
names(data)[51]<- "Y"
```

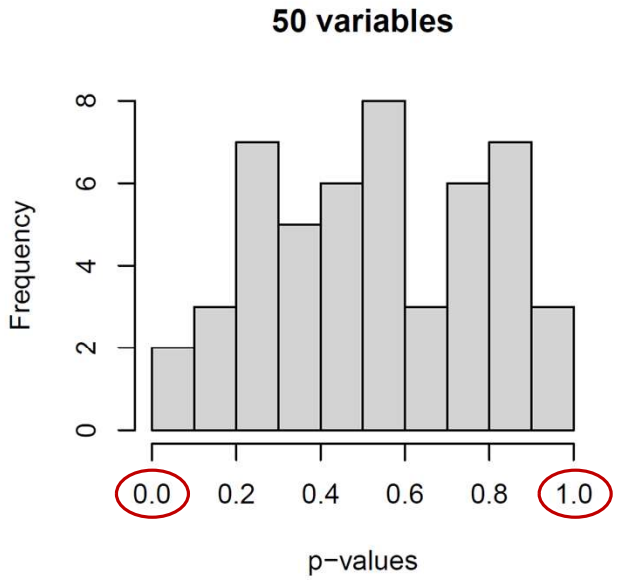
data is a 100$\times$ 51 matrix, where the last column is the response. The **data are generated randomly and independently**, the covariates do not have any explanatory power for the response!

2. Fit a linear model of y against all 50 variables

$$y_i = \beta_0 + \beta_1 x_i^{(1)} + \ldots + \beta_{50} x_i^{(50)} + \epsilon_i \ .$$

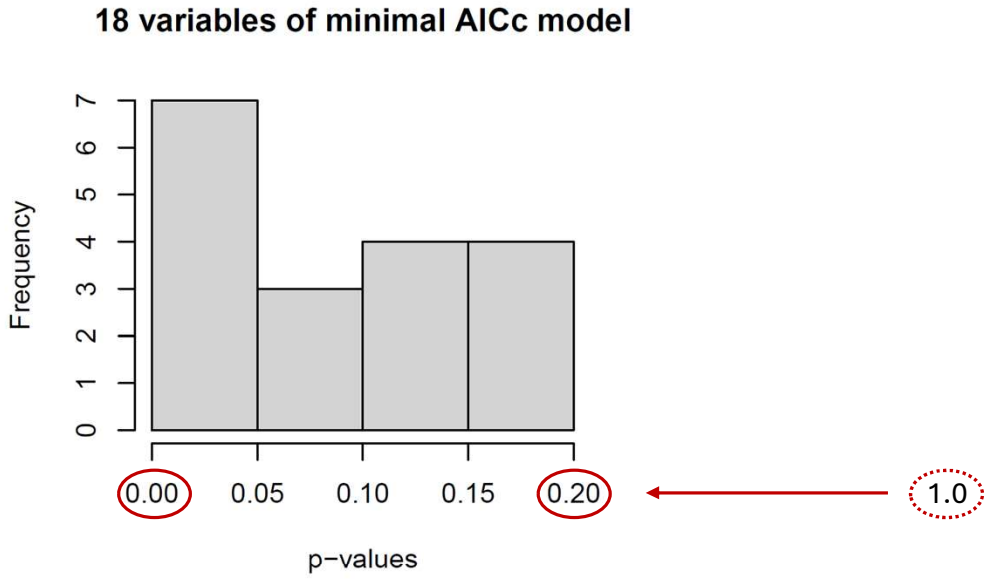
```
r.lm<- lm(Y~., data)
```

As expected, the distribution of the p-values is (more or less) uniform between 0 and 1, with none below 0.05:



3. Use AICc minimization to obtain the "best predictive" model:

```
r.AICmin<- stepAIC(r.lm, direction= "both",
                   trace= FALSE, AICc=TRUE)
```



The distribution of the p-values is now skewed: many reach rather small values, with 7 : $p < 0.05$. This happened although none of the variables has any explanatory power!

Type I errors (*i.e.* "false positives")

Main problem with model selection:

When model selection is carried out based on objective criteria, effect sizes will be too large, and the uncertainty too small → you end up being overly confident about effects that are too large.

Thus:

Model selection procedures *should not be used* when the aim of the analysis is explanation!

Variable selection using p-values?

In many publications you might see that people use p-values to select models instead. Also Stahel (Section 5.3) recommends this procedure.

However:

Variable selection using p-values is an especially bad idea.

$\rightarrow$ Please NEVER perform variable selection based on p-values$^{(\star)}$

What is the problem?

$^{(\star)}$Even not when the aim is prediction.

Importance is not reflected by p-values

A widely used practice to determine the "importance" of a term is to look at the p value ofthe t- or F-test and check whether it falls below a certain threshold (usually $p < 0.05$).

However, there are a few problems with this approach:

* **A small p-value does not necessarily imply that a term is (biologically, medically) important, or vice versa!**
* When carrying out the tests with $H_0 : \beta_j = 0$ for all variables sequentially, one runs into a multiple testing problem (Remember the ANOVA lecture of week 6, slide 25-26).
* The respective tests depend crucially on the correctness of the normality assumption.
* Variables are sometimes collinear, which leads to more uncertainty in the estimation of the respective regression parameters, and thus to larger p-values.

For all these reasons, we **disagree** with Stahel Section 5.2, second part in paragraph d.

> Statt die Tests für strikte statistische Schlüsse zu verwenden, begnügen wir uns damit, die P-Werte der t-Tests für die Koeffizienten (oder direkt die t-Werte) zu benützen, um die *relative* Wichtigkeit der entsprechenden Regressoren anzugeben, insbesondere um die „wichtigste“ oder die „unwichtigste“ zu ermitteln.

We also disagree with model selection based on p-values, as suggested in Section 5.3, because:

- This too will lead to model selection bias **Freedman 1983**.
- p-values are even less suitable for model selection than AIC/AICc/BIC for the reasons mentioned on the previous slide.

An explanatory model: Mercury example

The **research question** was:

"Gibt es einen Zusammenhang zwischen Quecksilber(Hg)-Bodenwerten von Wohnhäusern und der Hg-Belastung im Körper (Urin, Haar) der Bewohner?"

- *Hg concentration in urine* (Hg_{urine}) is the **response variable**.
- *Hg concentration in the soil* (Hg_{soil}) is the **explanatory variable** of interest.

In addition, the following variables were monitored for each person, as they might also influence mercury levels in a person's body:

smoking status; number of amalgam fillings; age; number of monthly fish meals; indicator if fish was eaten in the last 3 days; mother vs child; indicator if vegetables from garden are eaten; migration background; height; weight; BMI; sex; education level.

Thus: 13 additional (potentially confounding) **variables!**

How many variables can I include in my model?

Crude rule of thumb:

Include no more than $n/10$ (dummy) variables in your linear model, where n is the number of data points.

In the mercury example there are 156 individuals, so a **maximum of 15 (dummy) variables** can be included in the model.

Remarks: - Categorical variables with k levels require $k-1$ dummy variables. For example, if 'education level' has $k = 3$ categories, $k - 1 = 2$ parameters are used up.
- Whenever possible, the model should **not be blown up** unnecessarily. Even if there are many data points, the use of too many variables may lead to **overfitting**.
$\rightarrow$ See https://en.wikipedia.org/wiki/Overfitting.

In the mercury study, the following variables were included using *a priori* knowledge/expectations:

Variable	Meaning	type	transformation
Hg_urin	Hg conc. in urine (response)	continuous	log
Hg_soil	Hg conc. in the soil	continuous	log
vegetables	Eats vegetables from garden?	binary	
migration	Migration background	binary	
smoking	Smoking status	binary	
amalgam	No. of amalgam fillings	count	$\sqrt{\cdot}$
age	Age of participant	continuous	
fish	Number of fish meals/month	count	$\sqrt{\cdot}$
last_fish	Fish eaten in last 3 days?	binary	
mother	Mother or child?	binary	
mother:age	Interaction term	binary:continuous	

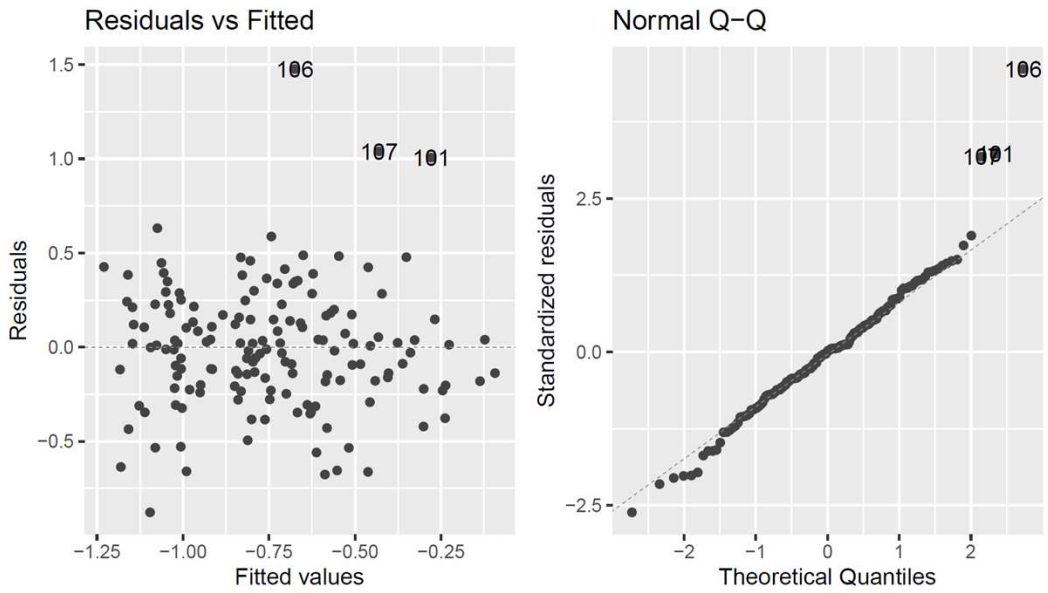
Let us now fit the full model in R:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.682	0.113	-6.057	0.000
log10(Hg_soil)	0.030	0.043	0.691	0.491
vegetables	0.058	0.057	1.019	0.310
migration	-0.013	0.087	-0.150	0.881
smoking	0.351	0.116	3.014	0.003
sqrt(amalgam)	0.290	0.052	5.616	0.000
age	-0.041	0.013	-3.289	0.001
mother	-1.087	0.373	-2.912	0.004
sqrt(fish)	0.070	0.030	2.352	0.020
last_fish	0.311	0.081	3.844	0.000
age:mother	0.057	0.016	3.627	0.000

You can interpret this table...
... but first validate the model!

- The p-value for mercury in soil, $\log_{10}(Hg_{soil})$, is rather high: p=0.49.

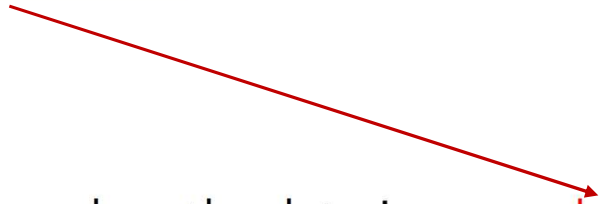
Always check the model, *e.g.* (see Lecture 5):



This looks ok, no need to improve the model from this point of view.

Once we've convinced ourselves the model can be trusted, we can ask questions like:

- Which of the terms in our model are **important/relevant**?
- Are there **additional terms** that might be relevant?
- Can we find **other patterns** in the data?



→ We could continue to analyse the data in an exploratory manner. Such additional models can be useful to generate new hypotheses.

For example, it might be tempting to check whether there are models with a lower AICc.

We could fit models from which certain terms are omitted. Let's try a model without the interaction *mother* · *age* (denoted as `r.lm0`).

```
AICc(r.lm0)
```

```
## [1] 135.0549
```

```
AICc(r.lm1)
```

```
## [1] 123.8577
```

The AICc increases quite a bit, confirming that the term is relevant.

In contrast, a model from which the binary *migration* variable is omitted, results in a reduced AICc:

```
r.lm0<- lm(log10(Hg_urin)~ log10(Hg_soil) + vegetables + smoking +
             sqrt(amalgam) + age * mother + sqrt(fish) + last_fish,
             d.hg)
AICc(r.lm0)
```

```
## [1] 121.5335
```

But: Given that the mercury model is an **explanatory, confirmatory model**, we should not remove a variable (*e.g.*, migration) simply because it reduces AICc.

→ Therefore, given the *a priori* selection of variables and the model validation results, the model from slide 34 was used in the final publication (Imo *et al.* 2017).

→ Any further analyses with other models needs to be labelled as exploratory.

Another complication: Collinearity

(See Stahel chapter 5.4)

Given a set of variables $x^{(1)}, x^{(2)}, x^{(3)}, ..., x^{(p)}$. If it is possible to write one of the variables as a linear combination of the others

$$x_i^{(j)} = \sum_{k \neq j} c_k x_i^{(k)} \quad \text{for all} \quad i = 1, ..., n$$

the set of variables is said to be collinear.

Examples: - Three vectors in a 2D-plane are always collinear. - Any variable that can be written as a linear combination of two others: $x^{(j)} = c_1 \cdot x^{(1)} + c_2 \cdot x^{(2)}$, then the three variables are collinear.

Basically: avoid including explanatory variables that are strongly correlated

In statistics, the expression "collinearity" is also used when such a collinearity relationship is *approximately* true. For example, when two variables $x^{(1)}$ and $x^{(2)}$ are highly correlated.

→ e.g. $r_{Pearson} > 0.7$

What is the problem with collinearity?

A simple (and extreme) example to understand the point: Assume two variables are identical $x^{(1)} = x^{(2)}$. In the regression model

$$y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \epsilon_i \ ,$$

the slope coefficients β_1 and β_2 cannot be uniquely determined (there are many equally "optimal" solutions to the equation)!

When the variables are collinear, this problem is less severe, but the β coefficients can be estimated less precisely

$\rightarrow$ standard errors too high.

$\rightarrow$ p-values too large.

Detecting collinearity

The Variance Inflation Factor (VIF) is a measure of collinearity. It is defined for each variable $x^{(j)}$ as:

$$VIF_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the R^2 of the regression of $x^{(j)}$ against all other variables (Note: if R_j^2 is large, this means large collinearity and thus a large VIF).

Examples

- $R_j^2 = 0 \rightarrow$ no collinearity $\rightarrow$ VIF=1/1 = 1.
- $R_j^2 = 0.5 \rightarrow$ some collinearity $\rightarrow$ VIF=1/(1-0.5) = 2.
- $R_j^2 = 0.9 \rightarrow$ high collinearity $\rightarrow$ VIF=1/(1-0.9) = 10.

(`vif()`-function in R)

What to do against collinearity

- Avoid it, *e.g.* in experiments.
- Do not include a variable with an unacceptably high R_j^2 or VIF_j. There are many critical VIF-values in the literature, ranging from 3 to 10.
- Be aware and interpret your results with appropriate care.
- See also Stahel 5.4(i) for a "recipe".

Note: We would probably not care much about collinearity in a predictive model. If collinearity was a problem, AIC/AICc/BIC would anyway select a subset where some collinearity is eliminated (because model complexity is balanced against model fit).

Recommended procedure for explanatory models I

Before you start:

- **Think about a suitable model**. This includes the model family (*e.g.*, a normal linear model), but also variables that are of interest, using **a priori** knowledge.
- Devise a strategy on how to handle when modelling assumptions are not met.
  - What kind of variable transformations would you try, in which order, and why?
  - What model simplifications will be considered if it is not possible to fit the intended model?
  - How to deal with outliers?
  - How to treat missing values in the data?
  - How to treat collinear variables?
  - ...

It is advisable to write your strategy down as a "protocol" before doing any analyses.

Recommended procedure for explanatory models II

Analyze the data following your "protocol":

- Fit the model and check if all assumptions are met.
- If assumptions are not met, **adapt the model** according to your protocol.
- Interpret model coefficients (effect sizes) and the p-values correctly (next week!).

Following the analysis that was specified in the "protocol":

- Any additional analyses that you did not specify in advance, are exploratory!

One more thing: Occam's Razor

This principle essentially states that an **explanatory model** should not be made more complicated than necessary.

This is also known as the principle of parsimony (Prinzip der Sparsamkeit):

> Systematic effects should be included in a model **only** if there is knowledge or convincing evidence for the need of them.

▸ See Wikipedia for "Ockham's Rasiermesser"

"Keep things as simple as possible, but not simpler"

Summary

- Model/variable selection is difficult and controversial.
- There are different approaches for predictive or explanatory models.
- Explanatory models are either confirmatory or exploratory.
- AIC, AIC_c, BIC: balance between model fit and model complexity.
- Automatic model selection leads to biased parameter estimates and p-values.
- Therefore, automatic model selection procedures are inappropriate for explanatory models.
- P-values should not be used for model selection, not even for predictive models.

References:

Brewer, M. J., A. Butler, and S.L. Cooksley (2016). The relative performance of *AIC*, AIC_c and *BIC* in the presence of unobserved heterogeneity. *Methods in Ecology and Evolution 7*, 679-692.

Burnham, K.P. and D.R. Anderson (2002). *Model selection and multimodel inference: a practical information-theoretic approach.* New York: Springer.

Clayton, D. and M. Hills (1993). *Statistical Models in Epidemiology.* Oxford: Oxford University Press.

Copas. J.B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 45*, 311-354.

Freedman, D.A. (1983). A note on screening regression equations. *The American Statistician 37,* 152-155.

Imo, D., S. Muff, R. Schierl, K. Byber, C. Hitzke, M. Bopp, M. Maggi, S. Bose-O'Reilly, L. Held, and H. Dressel (2017). Risk assessment for children and mothers in a mercury-contaminated area using human-biomonitoring and individual soil measurements: a cross-sectional study. *International Journal of Environmental Health Research 28,* 1-16.