CPUE standardization
using GLM

Outline

1. CPUE standartization – what is it and why do we need it
2. Methods for CPUE standardization
3. GLM
4. Model selection
5. Diagnostics

..with examples in R

CPUE standartization – what is it and why do we need it?

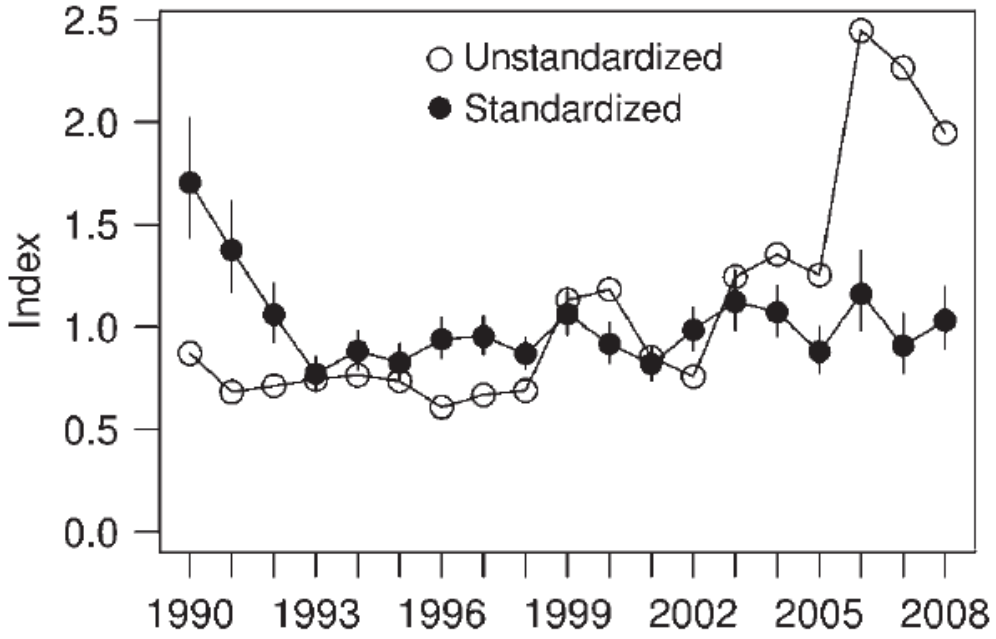
Removing variation in the data not attributable to changes in abundance.

The process of adjusting the time series of CPUE for the effects of other factors, such as:

- season (month),
- location/area/station,
- type of gear/gear material,
- vessel
- temperature, oxygen, wind, depth, etc.

CPUE standardization

- CPUE from commercial fishery



Bentley et al., 2012. ICES J Mar Sc, 69(1), 84–88. doi:10.1093/icesjms/fsr174

- CPUE from scientific surveys

Account for factors such as:

- Location/station
- Gear specifics (if different gears are used)
- Soak time (if varies)
- Environmental factors (wind, depth etc.)

Methods for CPUE standardization

- GLM
- GAM
- GLMM
- Regression Tree
- Neural network
- Delta Lognormal method
- …

Generalized linear model (GLM)

- Response variable can follow any distribution.
- does NOT assume a linear relationship between the response variable and the explanatory variables, but it does assume a linear relationship between the transformed expected response in terms of the link function and the explanatory variables
- The homogeneity of variance does NOT need to be satisfied
- Errors need to be independent but NOT normally distributed.

Generalized linear model (GLM)

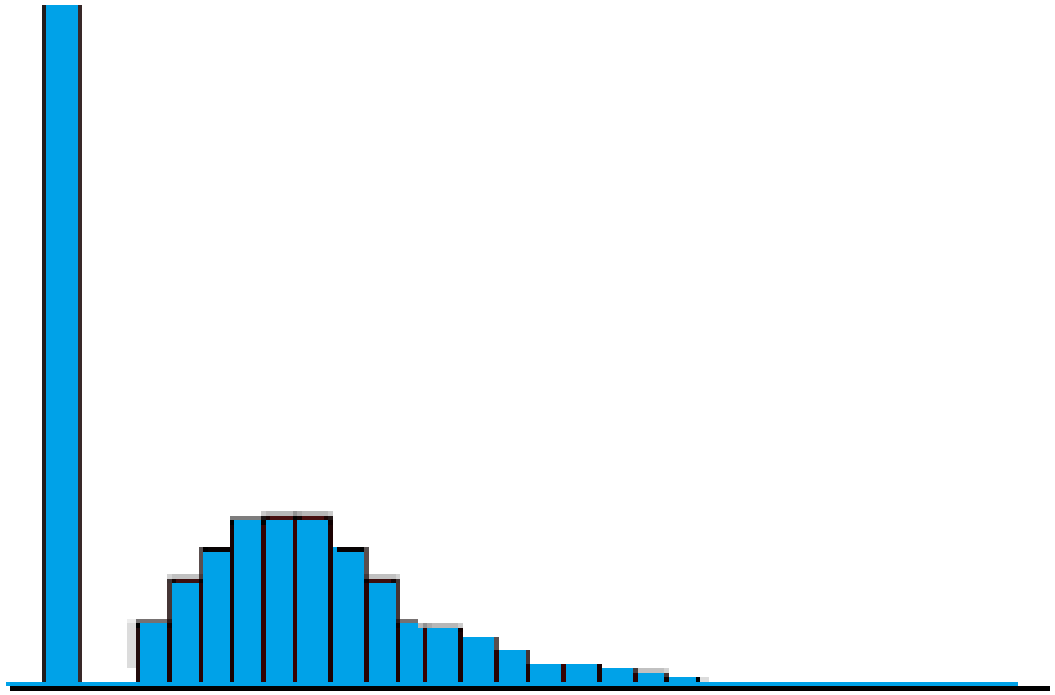
1. **Linear predictor -** a linear combination of parameter (*b*) and explanatory variable (*x*).
2. **Link function -** "links" the linear predictor and the parameter for probability distribution
3. **Probability distribution**

GLM for CPUE standartization. Steps

1. choose the response variable (e.g., **catch** per fishing event, in kg)
2. select a sampling distribution for the response variable (normal, exponential, Poisson, binomial, gamma, **tweedie,** etc.),
3. chose a link function appropriate to the distribution (e.g**., log link**)
4. select a set of explanatory variables (**year**, location, season, gear type, depth, etc.)

Tweedie distribution

- Have a cluster of data items at zero
- Very useful where we have mixture of zeros and non-negative data points.



Model selection. Which model is the best?

- Nested models - use anova, non-nested models – use AIC and BIC (Akaike information criterion and Bayesian information criterion)

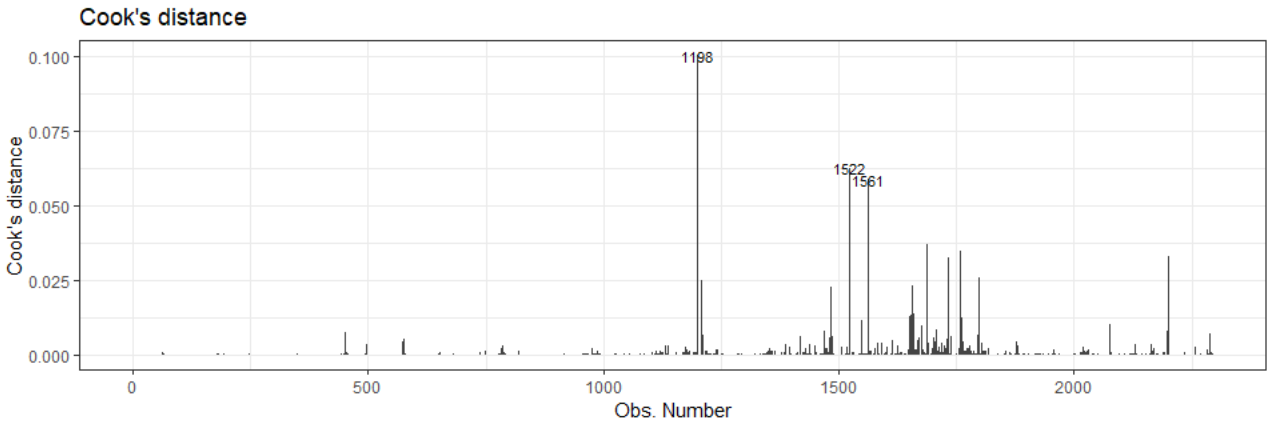
Steps:

1. Make a full model (which includes all factors)
2. Delete a term already in the model. Selection of terms for deletion is based on AIC (Akaike's information criterion). The procedure stops when the AIC cannot be improved.
3. Look into difference in explained deviance

Diagnostics

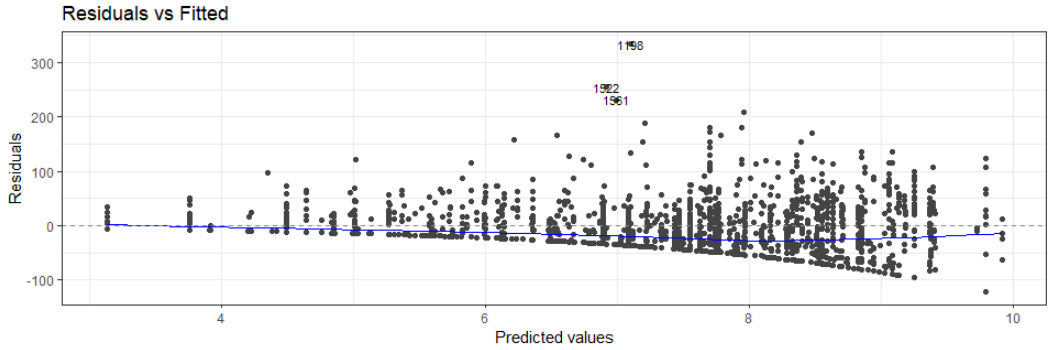
Outliers Detection (influence statistics)

- the *leverage* and the *Cook's distance*

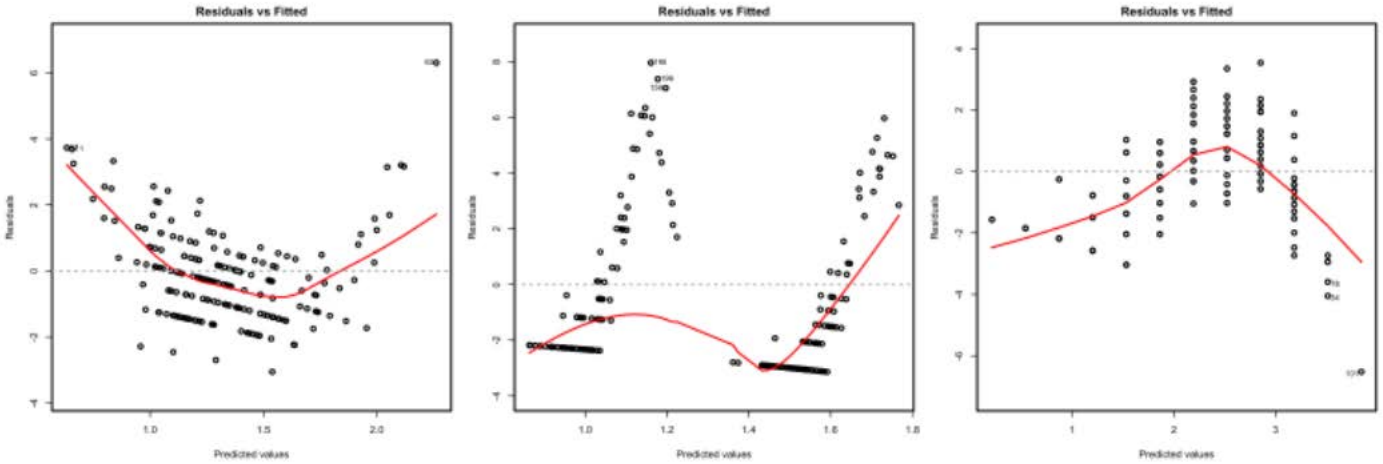


Diagnostics

Model assumptions - linearity



- The plot of (deviance) residuals against fitted values aims to check whether there are strong patterns.



Thank you!

www.sif.lt

Questions? Suggestions? Contact me:

Egle Jakubaviciute, PhD

egle.jakubaviciute@gamtc.lt

Researcher at

Laboratory of Fish Ecology

Nature Research Centre

Akademijos 2, 08412 Vilnius

2014-2020 Operational Programme for the European Union Funds Investments in Lithuania

This project has received funding from European Regional Development Fund (project No 01.2.2-LMT-K-718-02-0006) under grant agreement with the Research Council of Lithuania (LMTLT).