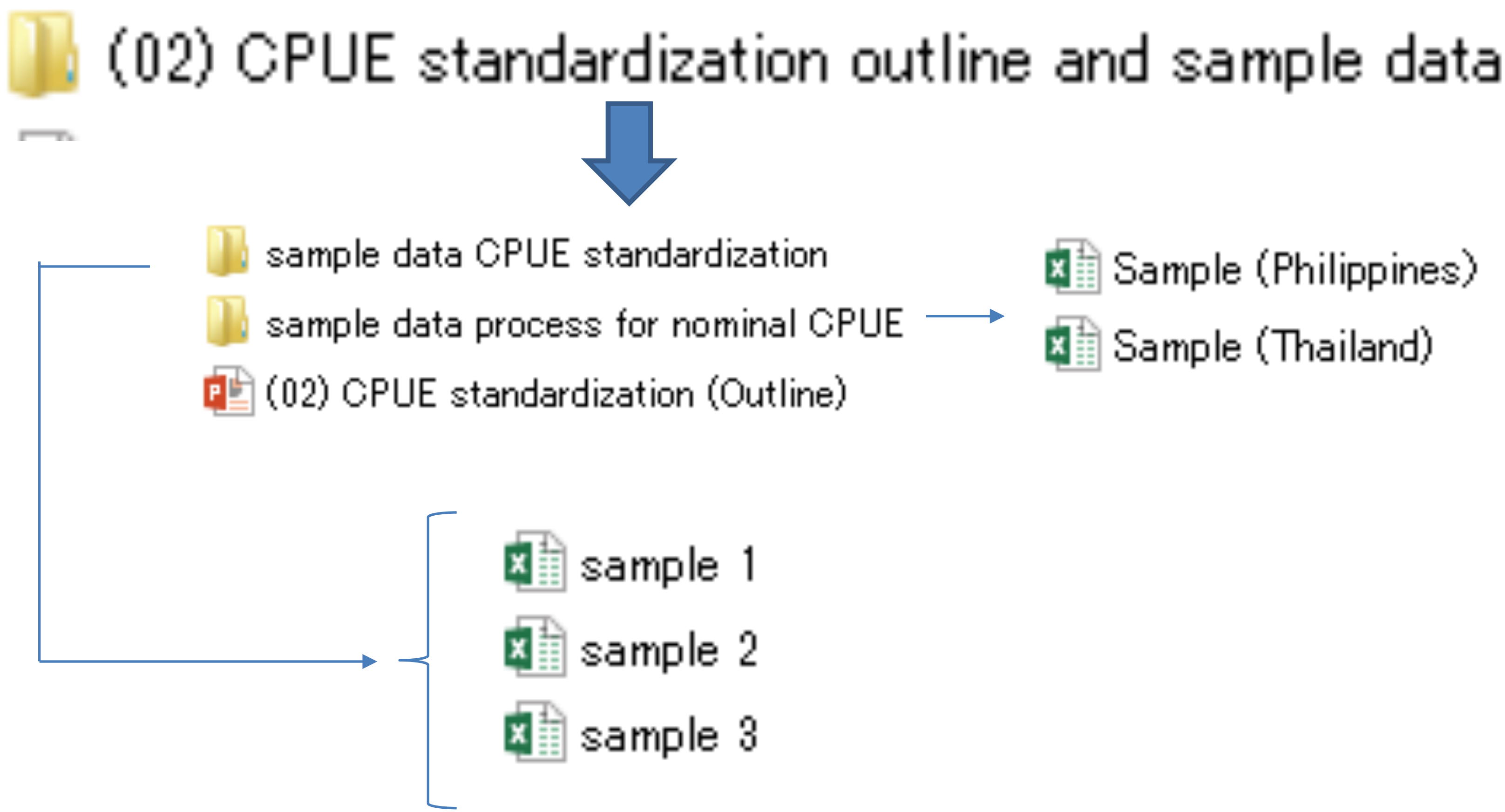
(02) CPUE standardization
April 18(Mon), 2016

- Before start, please copy



CPUE standardization
April 18(Mon)

3 major Agenda

OUTLINE

- Why do we need CPUE standardization?
  (Nominal vs standardized CPUE?）
- How to standardize CPUE ?
- What is our nominal CPUE?

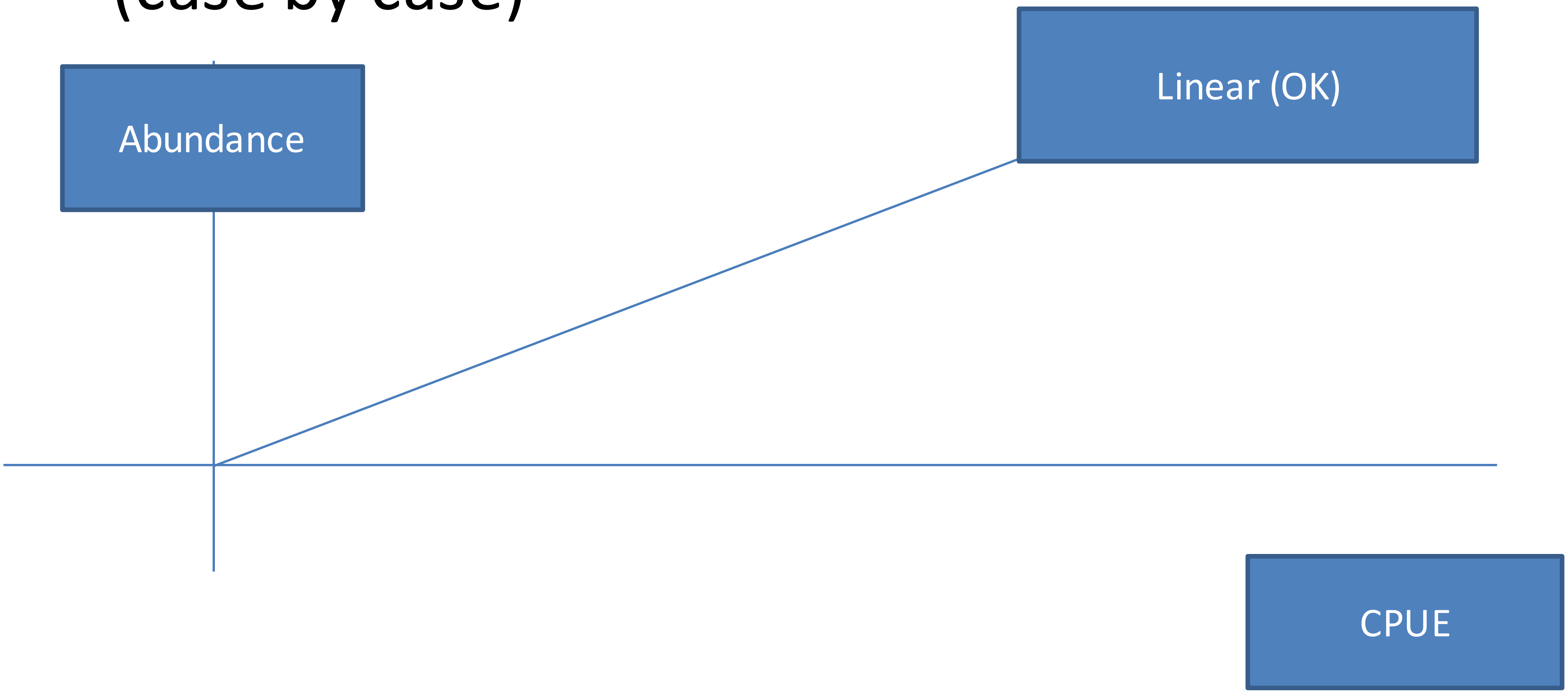
CPUE standardization by GLM (soft)

- (03) CPUE standardization software by GLM(manual)
- Installing software
- Practice using the GLM software by participants

Data processing to create Nominal CPUE (Excel)

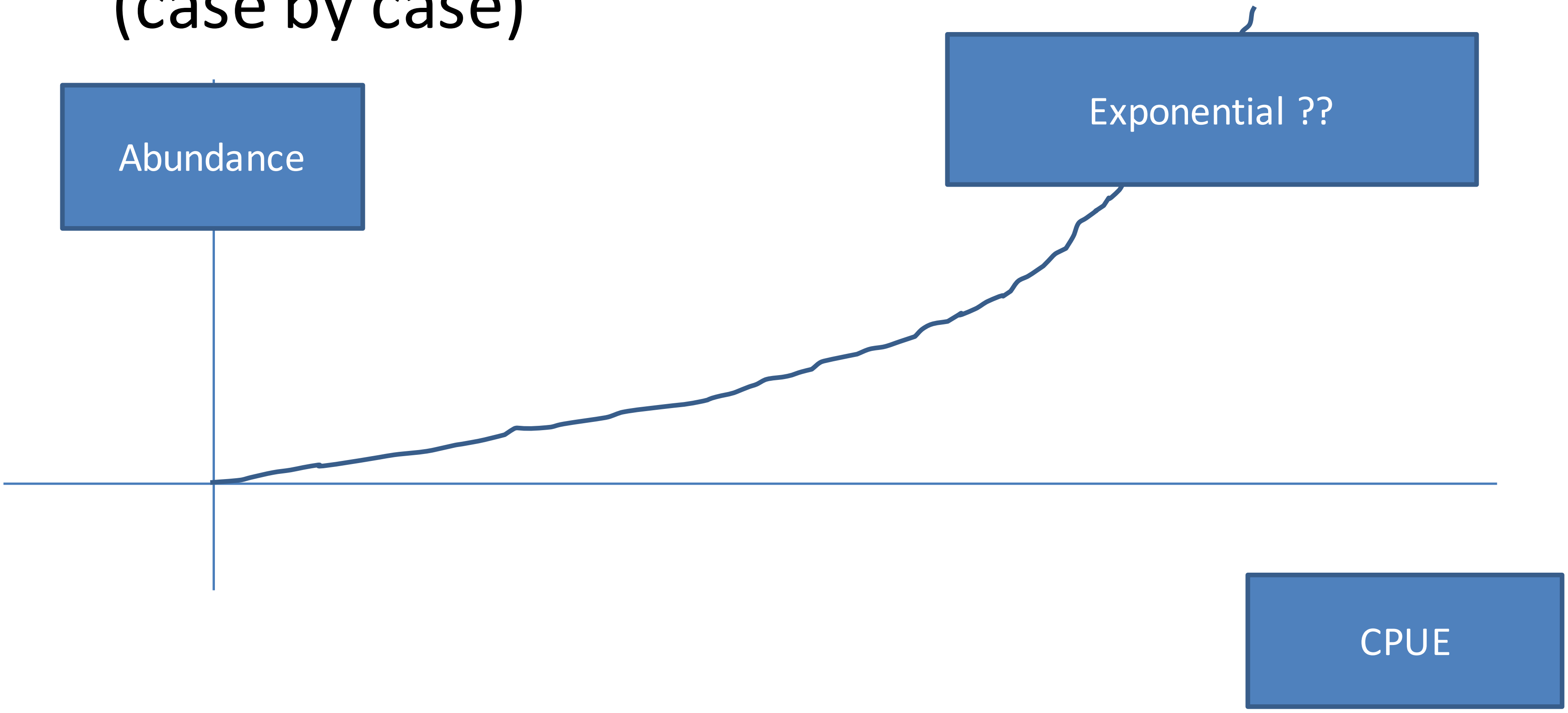
What is CPUE?

- Catch/fishing effort ➔ catch rate
- Approximation of abundance ?   Yes or no?
  (case by case)



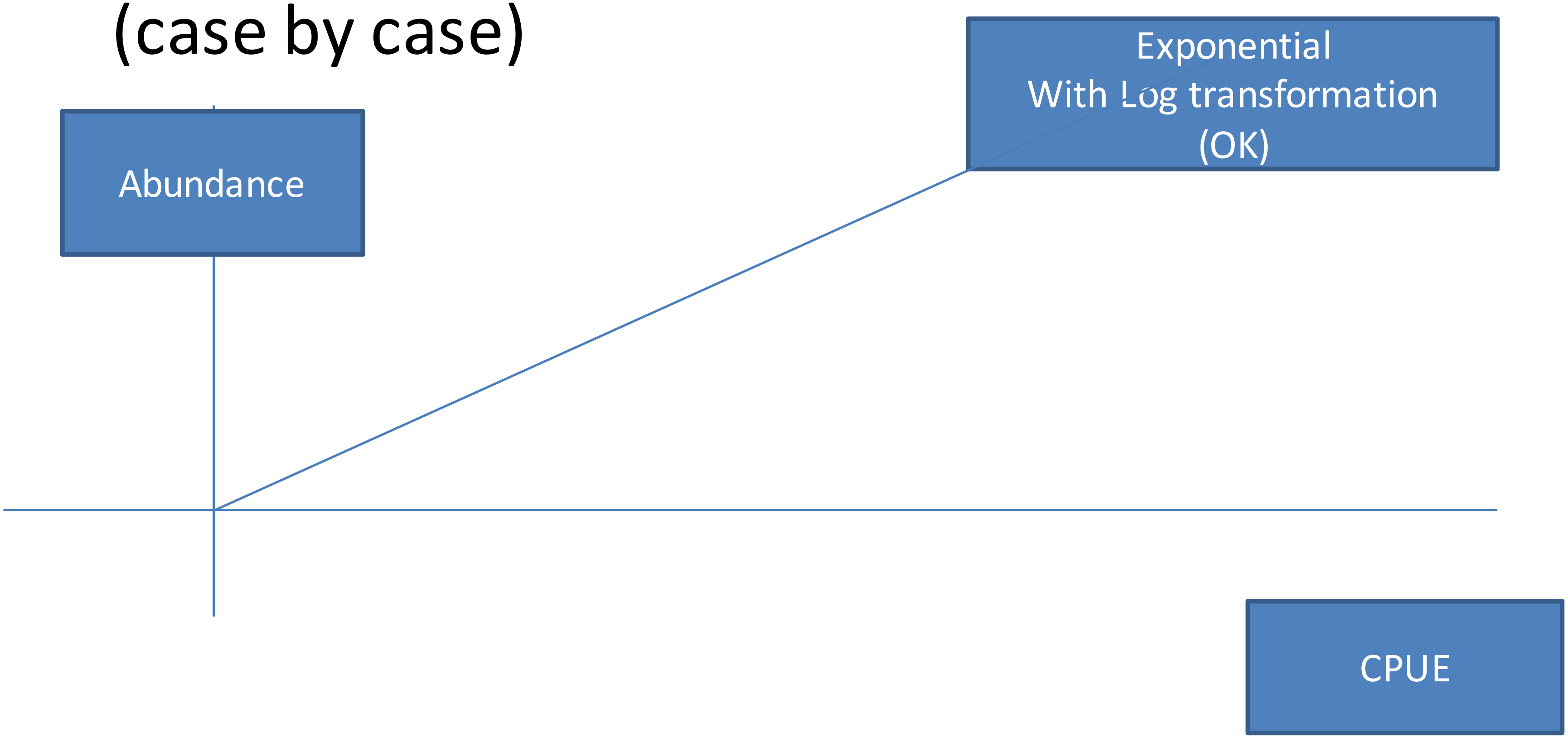
What is CPUE?

- Catch/fishing effort ➔ catch rate
- Approximation of abundance ? Yes or no? (case by case)



What is CPUE?

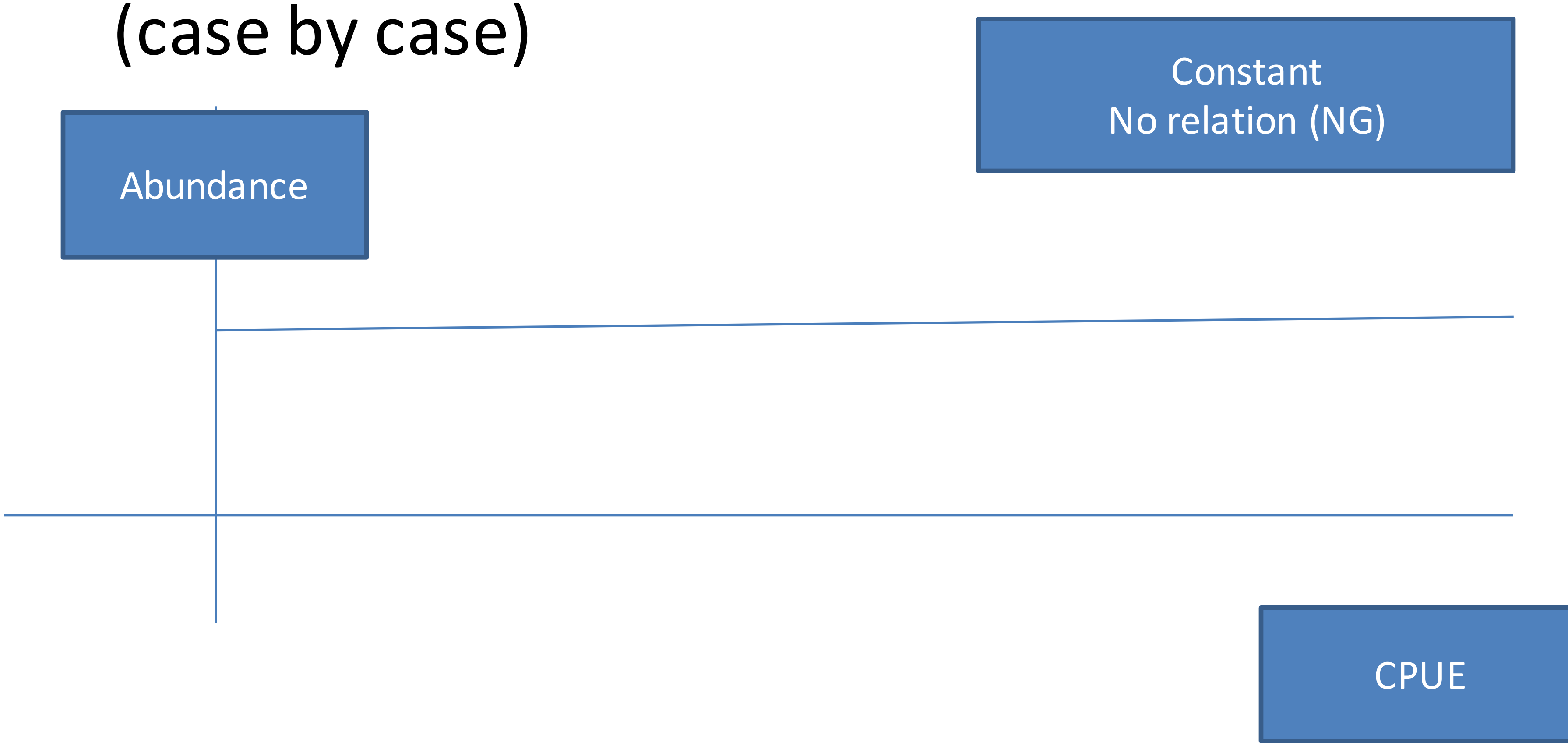
- Catch/fishing effort ➔ catch rate
- Approximation of abundance ? Yes or no?
  (case by case)



What is CPUE?

- Catch/fishing effort → catch rate
- Approximation of abundance ? Yes or no?

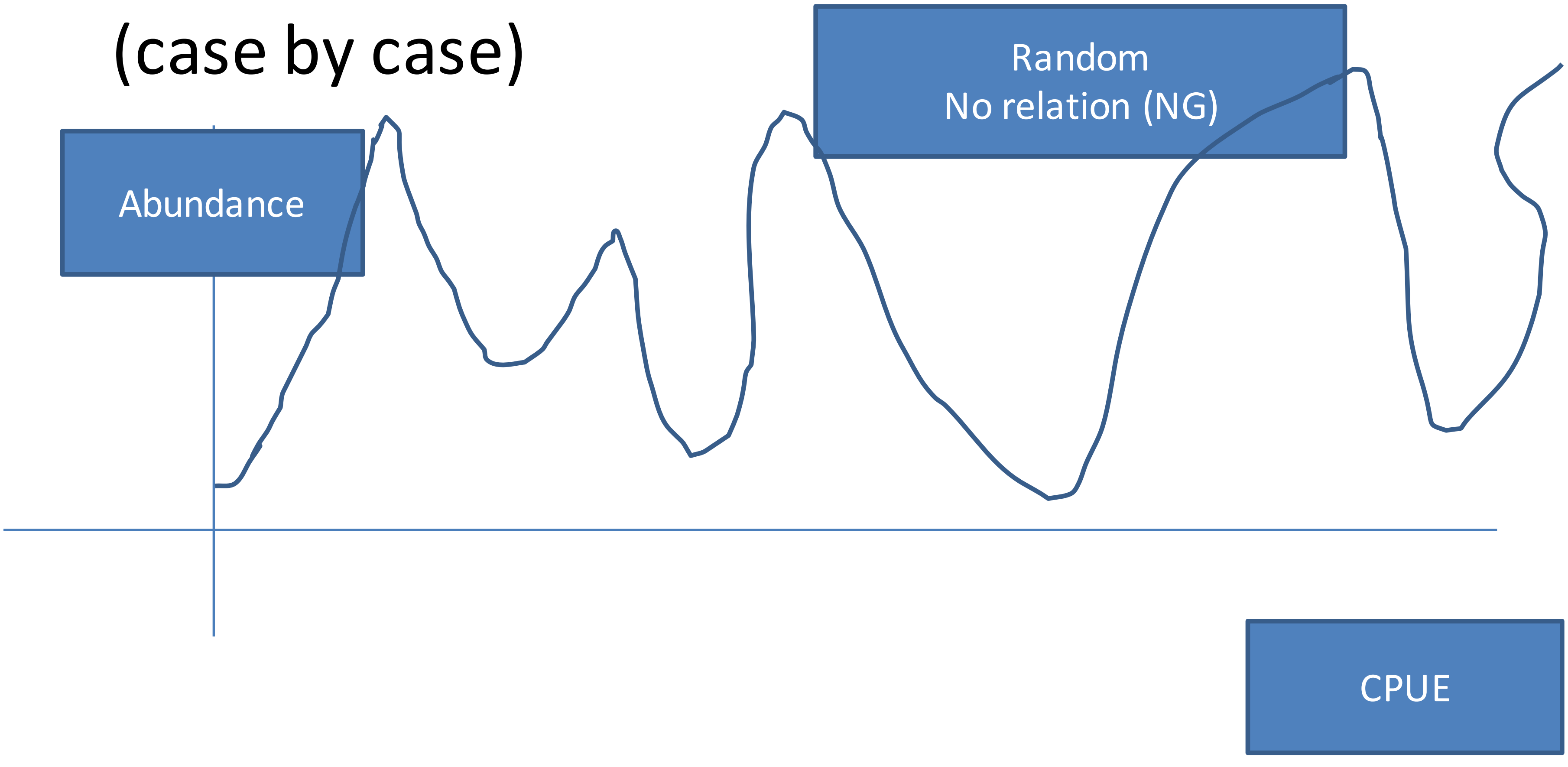
  (case by case)



What is CPUE?

- Catch/fishing effort ➔ catch rate
- Approximation of abundance ? Yes or no?

(case by case)



CPUE is proxy of the abundance

- Yes if CPUE can represents the linear relation to the abundance (with transformation)

- There are a number of transformation (equations)

Why do we need to standardize nominal CPUE? 2 reasons

(1) To filter biases in N_CPUE

Nominal CPUE : raw data [Catch/Effort]

↑

Many factors affect nominal CPUE

e.g., Year, season, area, ENV, evolution of gears..

Gear configuration (targeting)

→ **Produce BIAS**

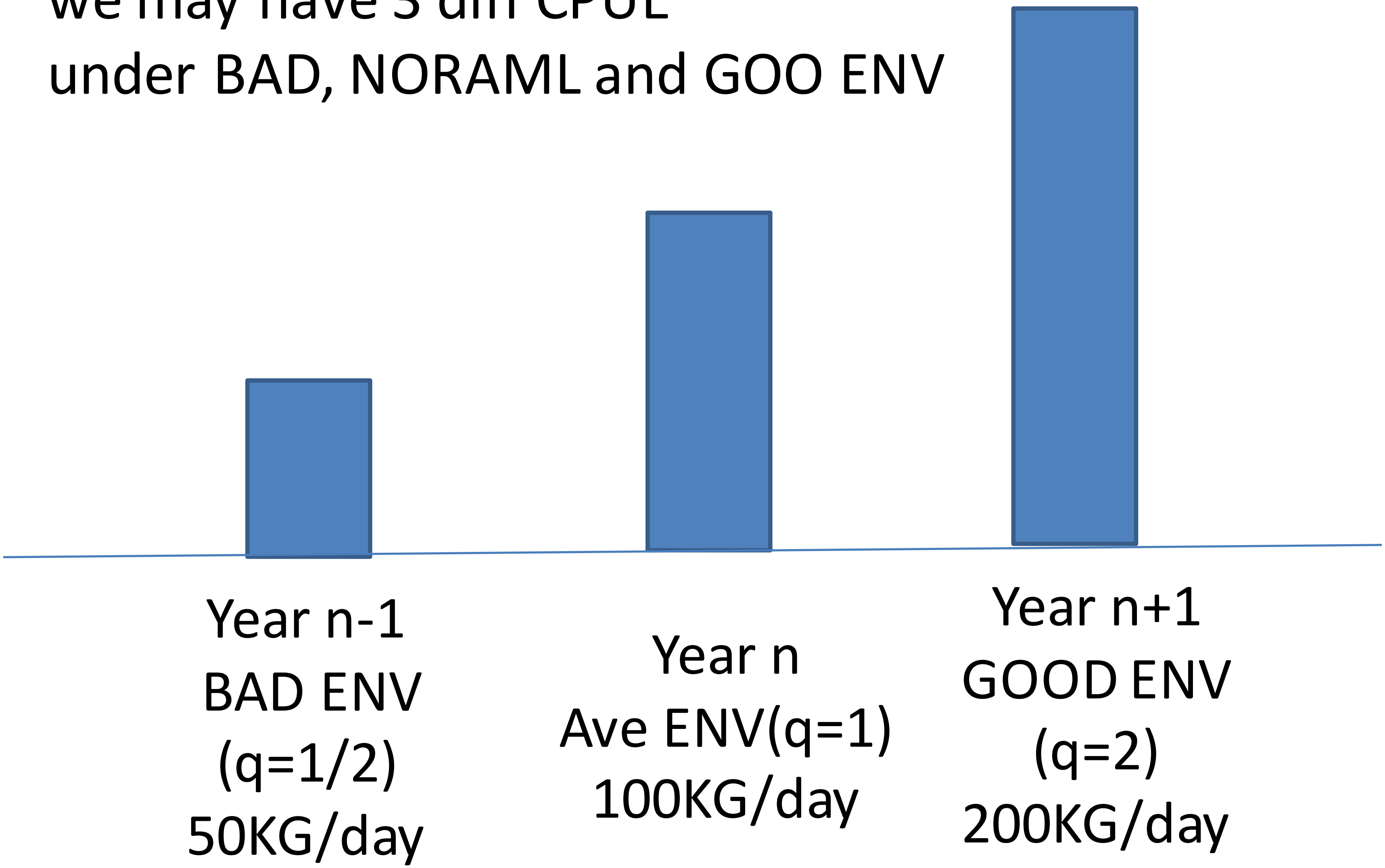
(2) To learn relation between CPUE vs. abundance

Simple example **RE: biases**

same area in 3 different years

we may have 3 diff CPUE

under BAD, NORAML and GOO ENV



q

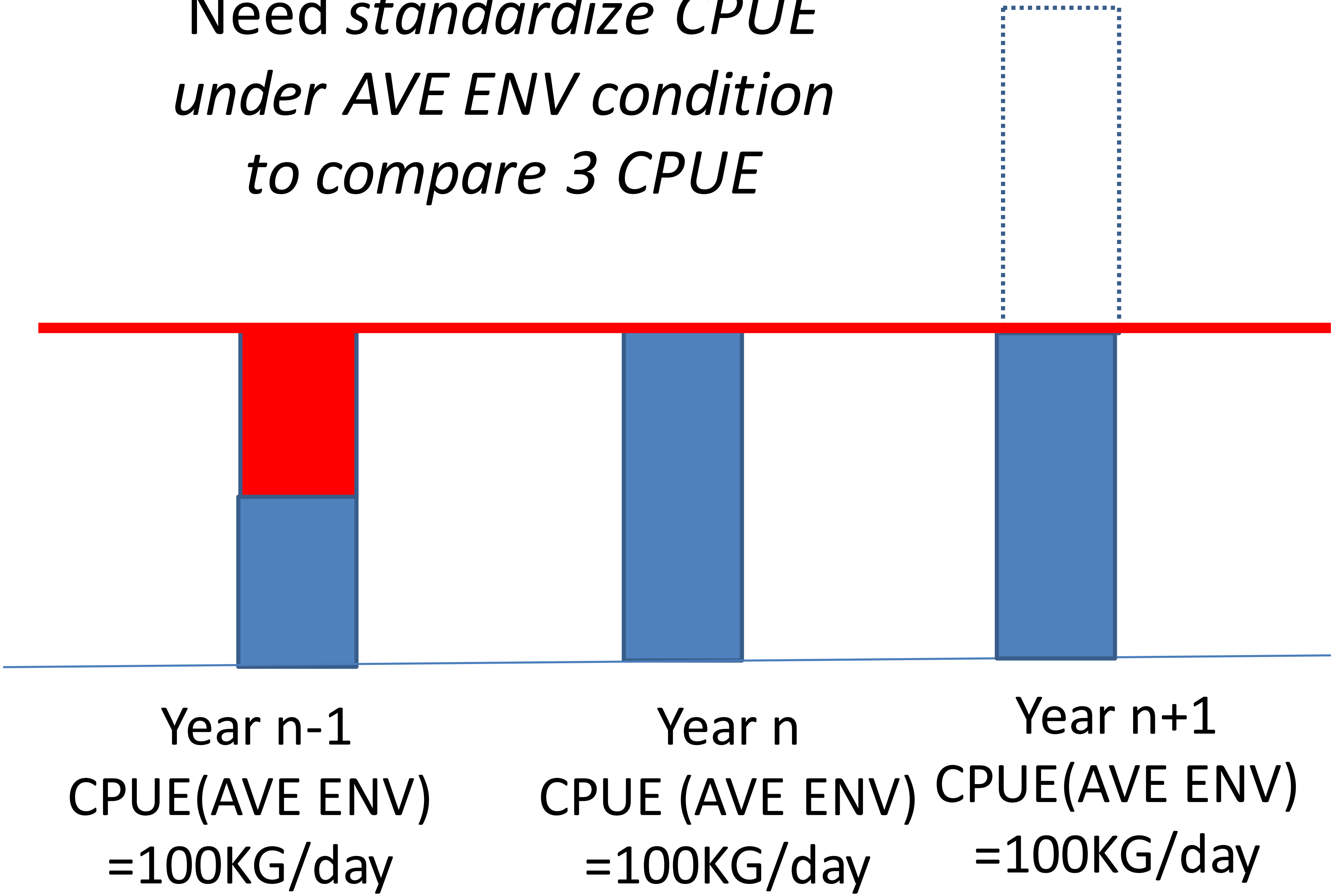
Catchability : efficiency of catch

Boat A q= 1   vs. Boat B q= 2

Boat B can make 2 times of catch in the same fishing effort

Boat B 2 time efficiency

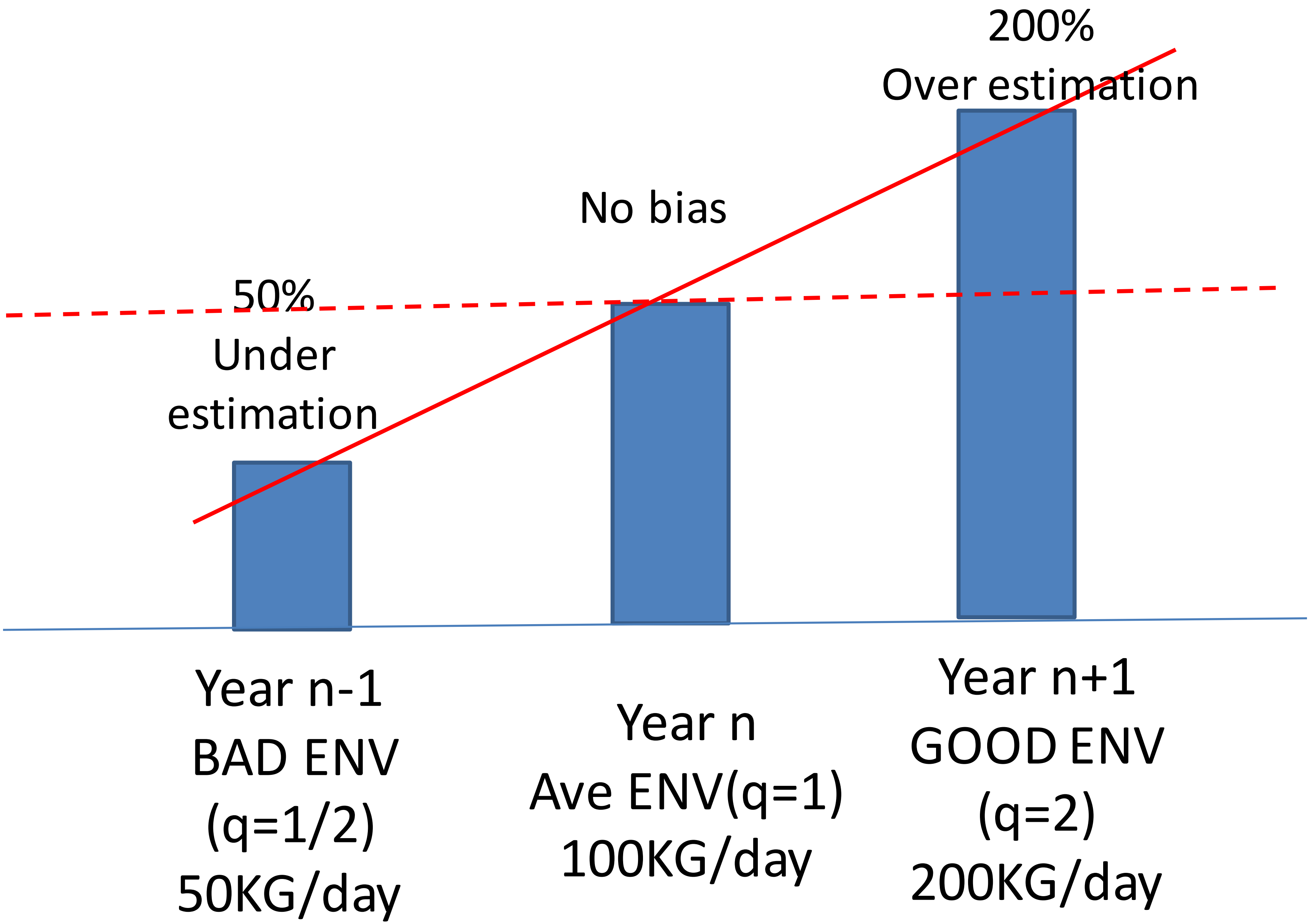
Need *standardize CPUE under AVE ENV condition to compare 3 CPUE*



After standardizing CPUE

3 year of CPUE (n-1, n and n+1) are actually same level

If you don't standardize CPUE you will have biases



But reality is very complicated many factors affect nominal CPUE

Major factors affecting nominal CPUE

Year, Season and Area

There are many other factors

ENV, Skipper ability, evolution of fishing gear etc..

But reality is very complicated many factors affect nominal CPUE

For this time we consider these 3 major factors
Year, Season and Area
as a first step....

Actually these 3 factors explain major biases based on many case studies

Even you add other factors their contribution are minor

How to proceed
CPUE standardization

For this bias adjustment (i.e., CPUE standardization) we use the statistical method

GLM: Generalized Linear Model

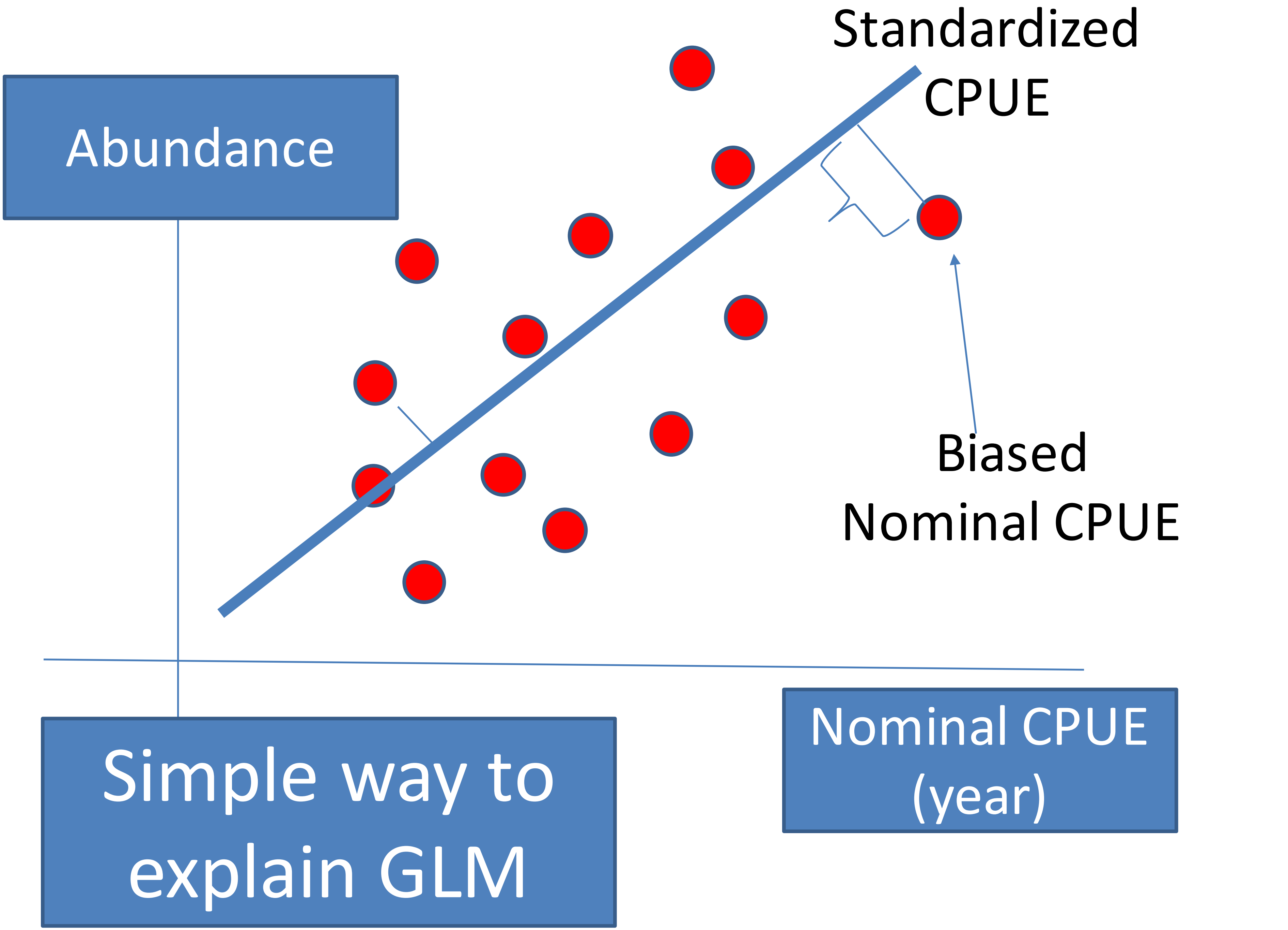
This is like simple regress

There are many other methods

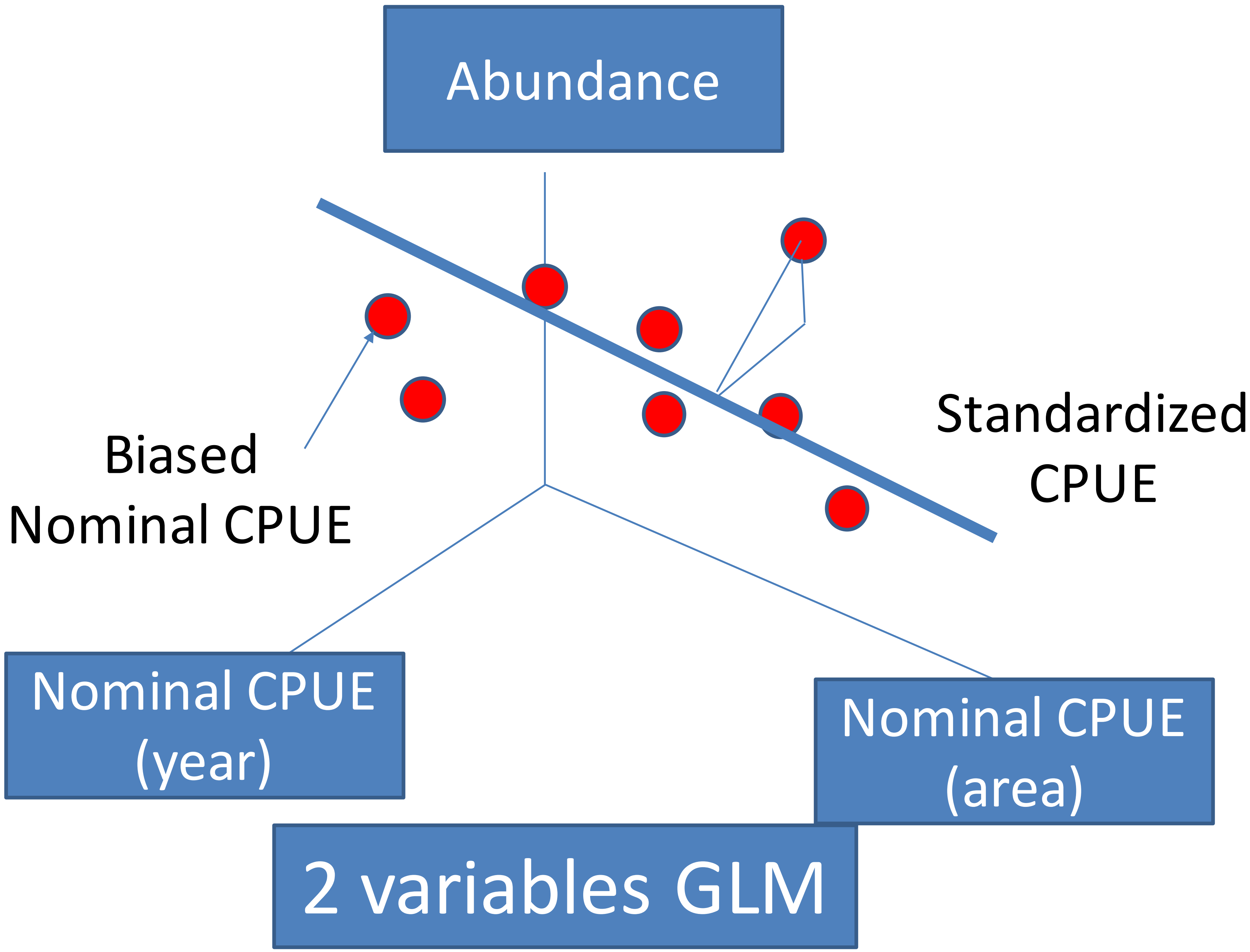
for different data (error) structure..

(GAM, binominal model, regression tree..)
(future)(beyond our scope)

But the GLM is the standard approach


Abundance
Standardized CPUE
Biased Nominal CPUE
Nominal CPUE (year)
Simple way to explain GLM

If you have more than 2 variables (multivariate situation) you have more complex way to compute to standardize CPUE


Abundance
Biased
Nominal CPUE
Standardized
CPUE
Nominal CPUE
(year)
Nominal CPUE
(area)
2 variables GLM

What happen if 3 or more variables ?

- Can not draw images …

- More complicated to compute
  So we need statistical package Software

statistical package Software multivariate statistical analyses

• SAS	$$(high)
• SPSS	$$ (medium)
• BMDP	$$ (medium)
• R	free (Flexible)
• Statistica	$$ (low)
• Excel	$$ (low) difficult to do

Our GLM model

Log_(N_CPUE+ constant)

= (mean)+[yr]+[Season:Q]+ [Area] + error

constant : 10% of N_CPUE

(Nishida & Campbell, 1998)

To avoid the log(0) problem without affecting the N_CPUE

Area

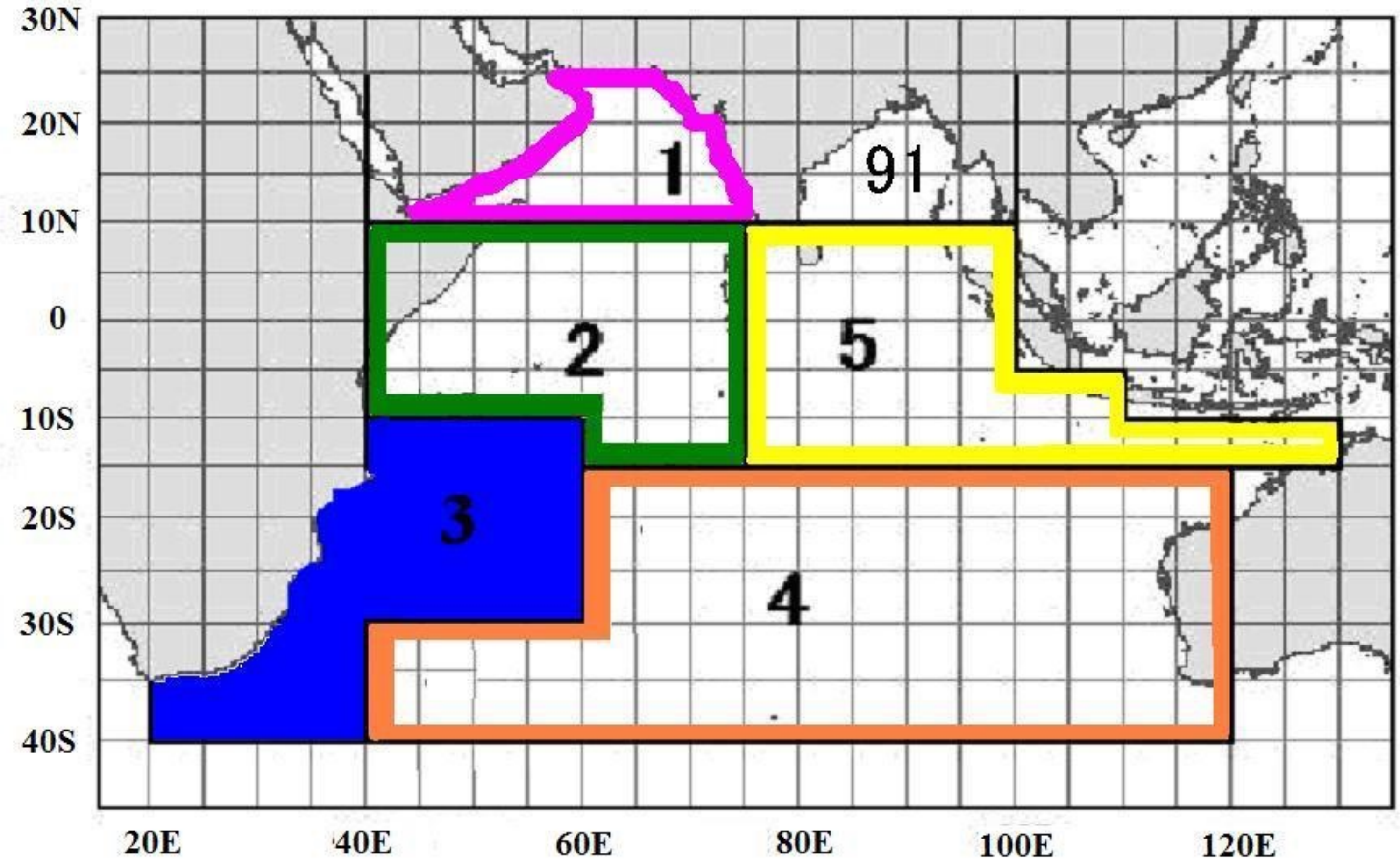
Fishing area : artisanal fisheries

➔ statistical area (Polygon) (see example)

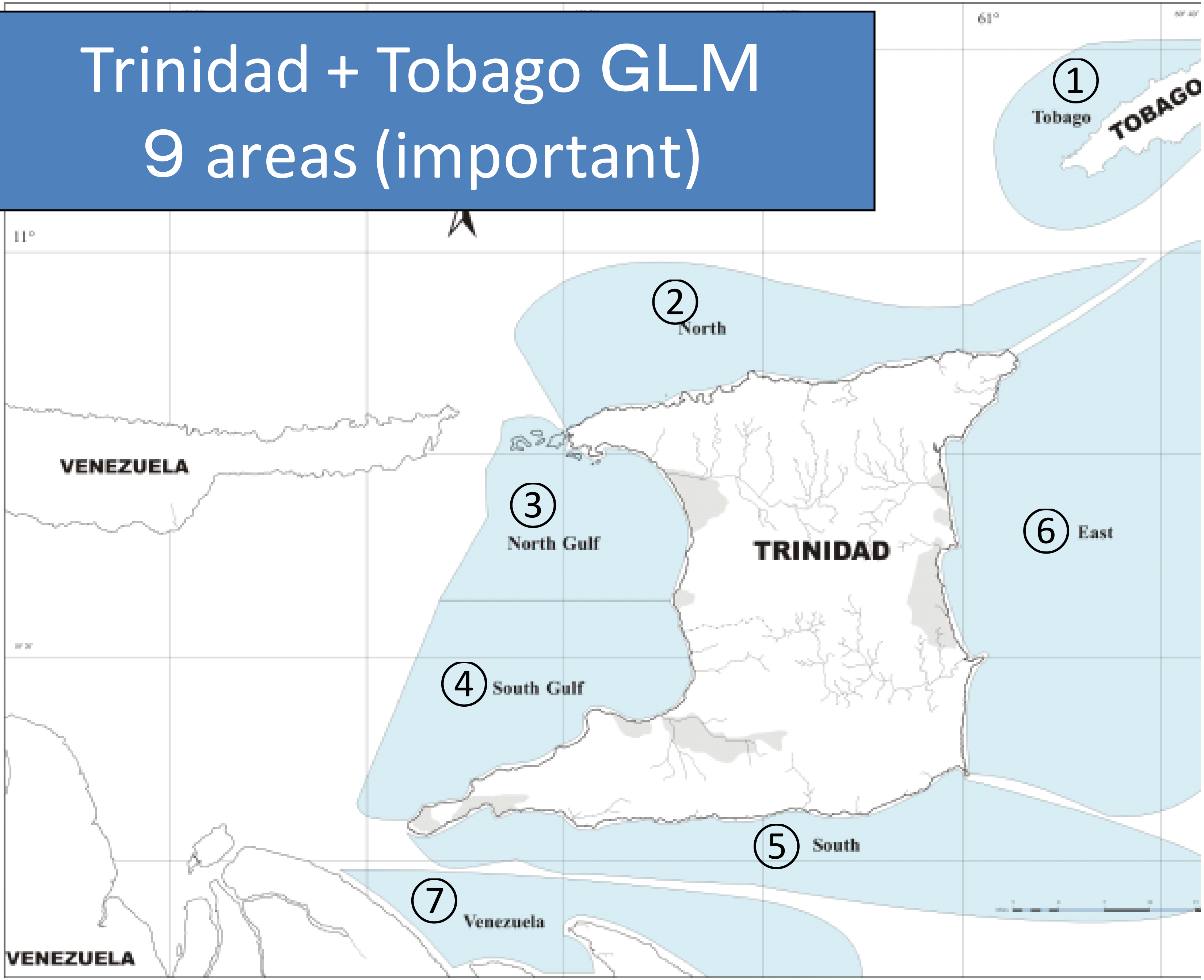
➔ Tuna longline

Grid 1x1 or 5x5 area (log book)

YFT (tuna LL)



Trinidad + Tobago GLM
9 areas (important)



Evaluation of STD_CPUE vs. global Catch

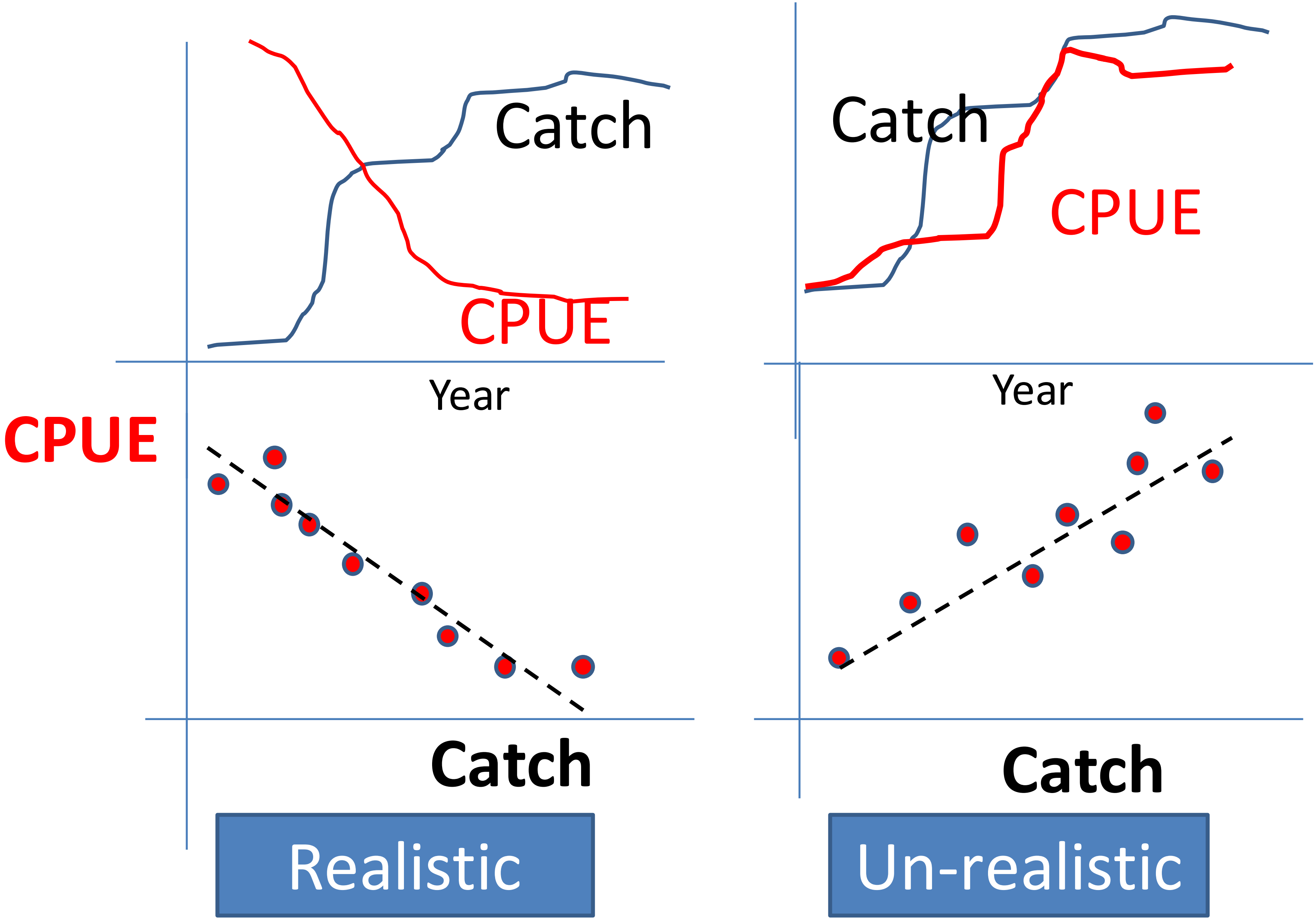
- Important to evaluated realistic CPUE

(1) Check Sample size

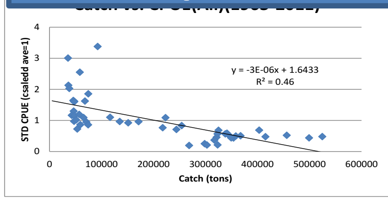
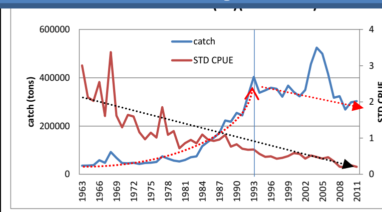
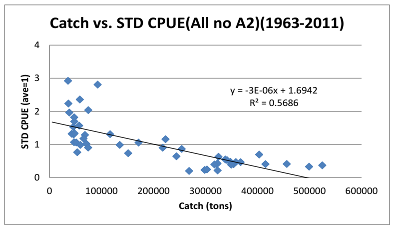
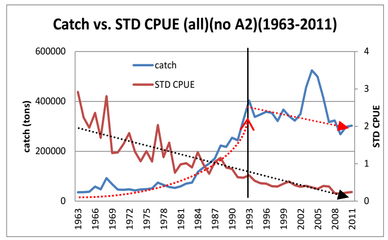
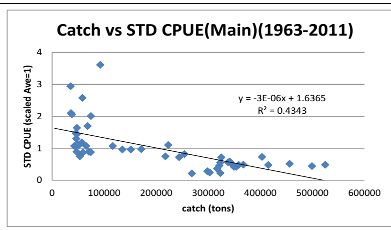
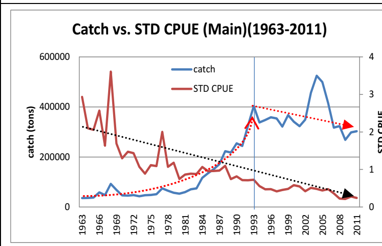
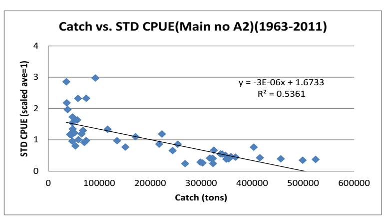
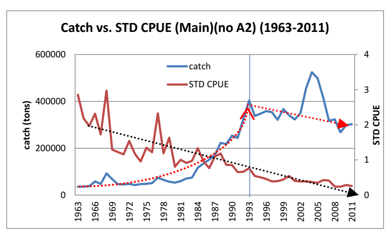
Gear , boat, area, month (season)

Select with large sample size

(2) Check relation Catch vs. CPUE => inverse CORR (realistic)

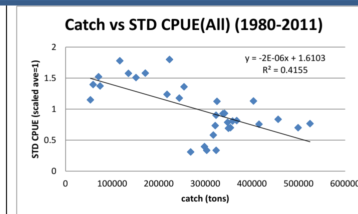
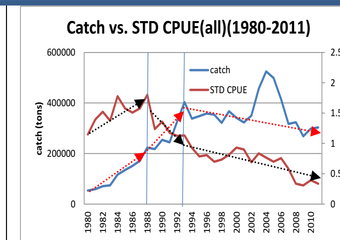
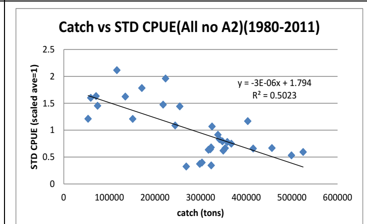
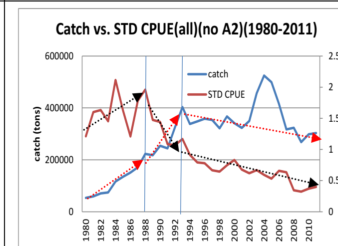
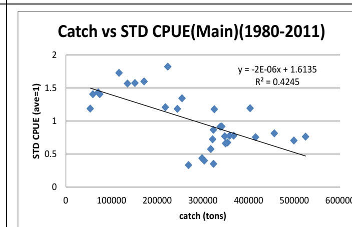
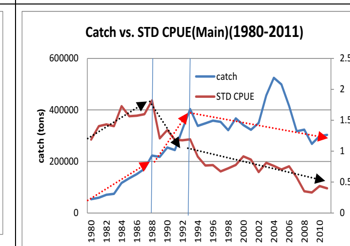
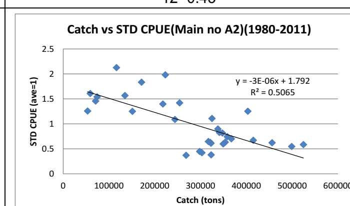
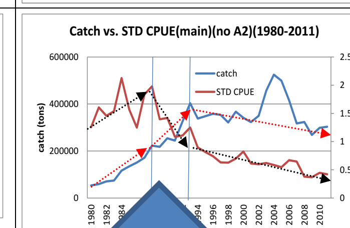


Long (1963-2011) (ave: r2=0.52)

No		
(1) All 2 3 4 5	 r2=0.46	
(2) All 3 4 5	 r2=0.57	
(3) Main 2 3 5	 r2=0.48	
(4) Main 5	 r2=0.59 (Best)	
feature	*Generally there are negative (realistic) correlations between catch vs. STD CPUE (ave r2=0.52)*	*There are 2 period … relations between catch vs. STD CPUE, i.e … st one (1980-1993) for negative (rea … ns and the 2nd one (1993-2011) for p … nrealistic).*

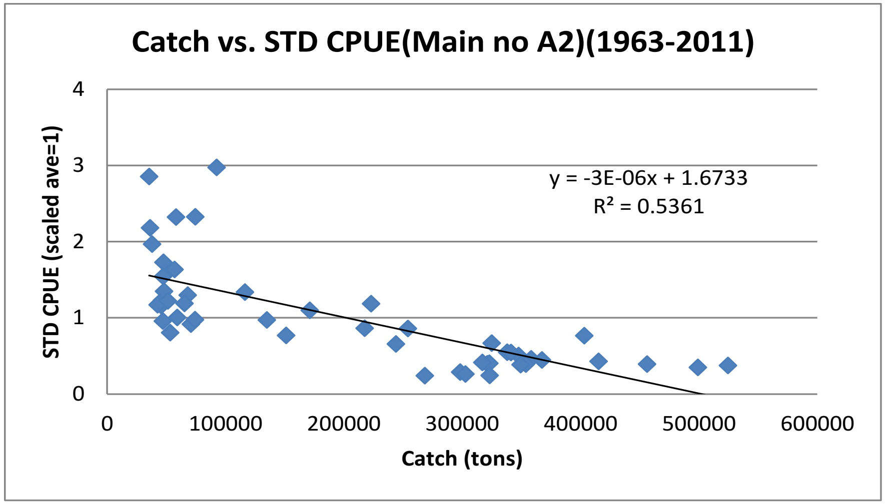
O K

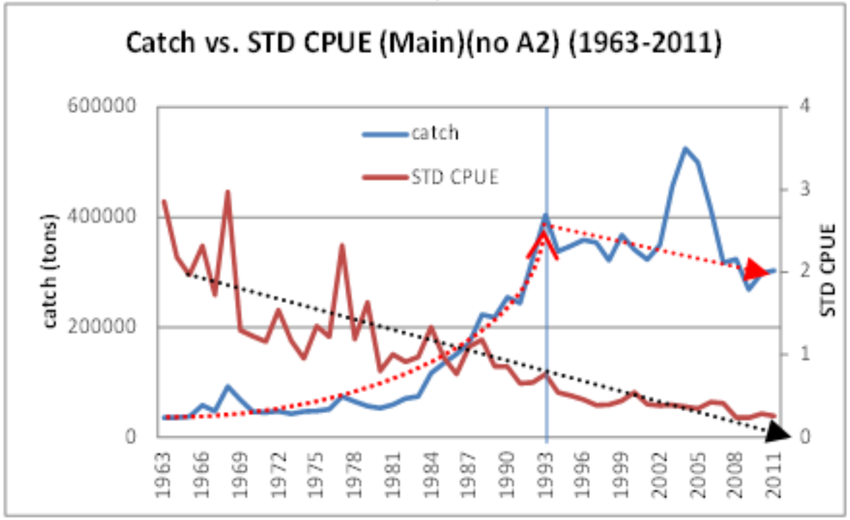
Short (1980-2011) (ave: r2=0.46)

(1) All 2 3 4 5	 r2=0.42	
(2) All 3 4 5	 r2=0.50	
(3) Main 2 3 5	 r2=0.43	
(4) Main 3 5	 r2=0.51	
	Generally there are negative (realistic) correlations between catch vs. STD CPUE (ave r2=0.46)	*There are 3 o … iods in relations between catch vs. STD … ., 2 periods (1980-88 and 1993-2011) sh … e (unrealistic) correlations and the other … 92) for negative (realistic)*

O K

Selected STD CPUE (Main no area 2)





How to proceed the GLM

First we need to make nominal CPUE data set
(we will practice later)

Simple example

year	Q	area	catch(kg)	effort (haul)	CPUE
1995	1	NC	45	9	5 (kg/haul)

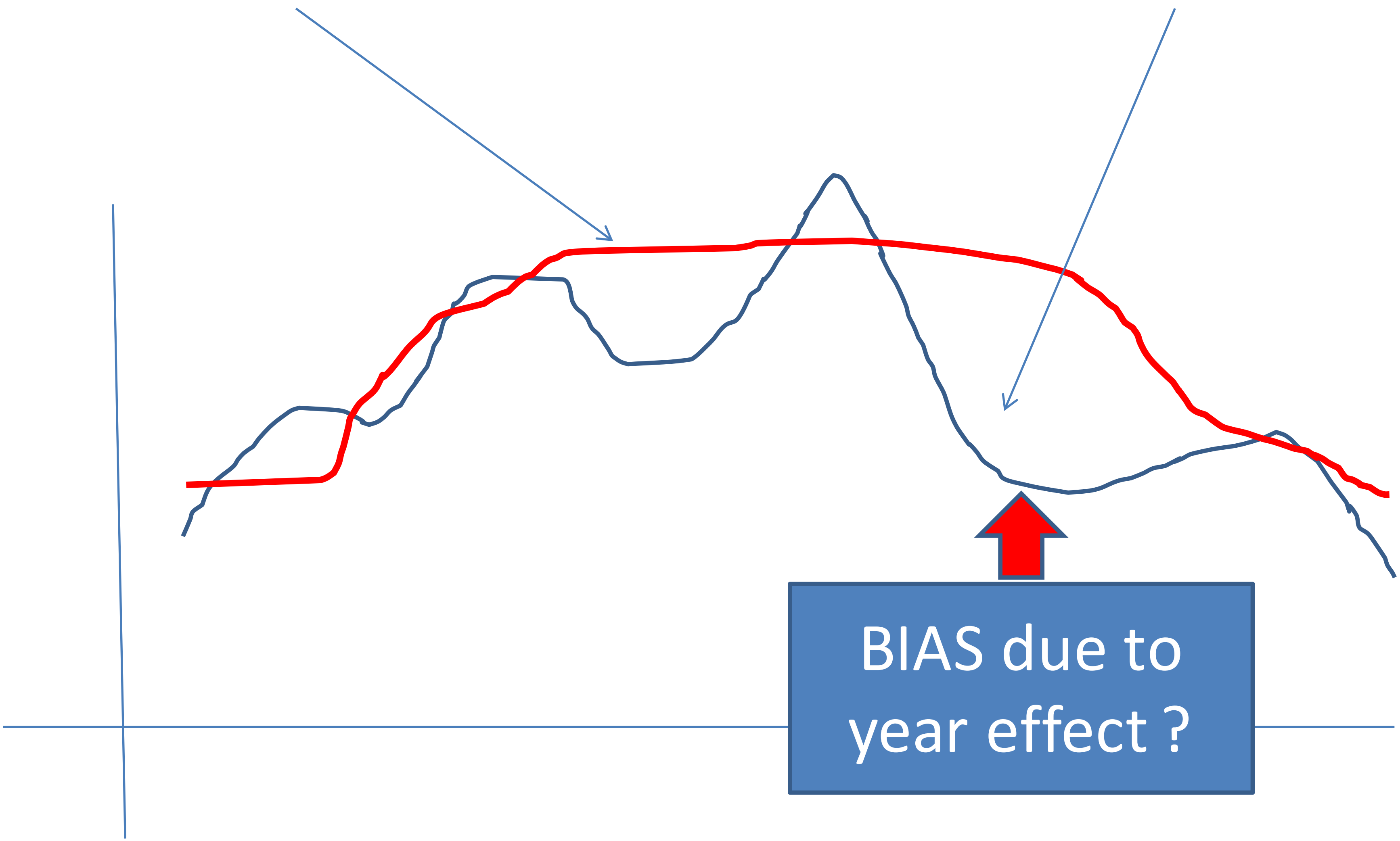
Data set (simple sample)

YR	Q	area	KAW CPUE (KG/HAUL)
2006	1	6	26.88
2006	1	6	0.00
2006	1	6	0.00
2006	2	6	163.35
2006	2	6	314.64
2006	2	6	37.69
2006	3	6	237.87
2006	3	6	429.18
2006	3	6	18.69
2006	4	6	29.62
2007	1	6	0.00
2007	1	6	0.00

Some example…(more complicated) (Japan tuna LL: Swordfish N_CPUE)

yr	hpb	mo	day	lat	ns	long	ew	hook	swo_n	area	swo_w	eda	v	miki	a	DPI	IOI
1990	13	12	2	26	2	40	1	2520	0	89	0	2	KASUGA MARU NO.18	2	7	-0.38	-0.68
1990	13	12	3	26	2	40	1	2520	0	89	0	2	KASUGA MARU NO.18	2	7	-0.38	-0.68
1990	13	12	4	26	2	40	1	2520	0	89	0	2	KASUGA MARU NO.18	2	7	-0.38	-0.68
1990	7	12	4	27	2	39	1	2660	0	89	0	2	KASUGA MARU NO.88	2	7	-0.38	-0.68
1990	7	12	5	27	2	39	1	2660	0	89	0	2	KASUGA MARU NO.88	2	7	-0.38	-0.68
1990	7	12	9	27	2	39	1	2660	0	89	0	2	KASUGA MARU NO.88	2	7	-0.38	-0.68

STD CPUE smooth out N_CPUE



Installing the software

Please refer to

(3) Manual of CPUE standardization by GLM

SPECIFICATION 1

- This software was developed using "R computer language" to perform nominal CPUE (Catch-Per-Unit-of-Effort) standardization by GLM (Generalized Linear Model) from fisheries data.

SPECIFICATION 2

- Nominal CPUE means Catch/Fishing effort, where catch and fishing effort data in the original (raw) data.
- GLM is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution.

SPECIFICATION 3

- This software can conduct “GLM” and create relevant outputs using menus without making any programming.
- Main input (covariates) are “Year” effect, “season” effect, “area” effect. As for outputs, standard GLM outputs created by “R language” will be automatically produced for each GLM run.

SPECIFICATION 4

- Major outputs are (a) ANOVA table, (b) Coefficients of estimated parameters, (c) Estimated annual standardized CPUE, (d) residual plots and (e) Q-Q plots.

Important note

This soft was just made last week (1st version)

Covariates (year, season and area) by GLM

In the future

with interaction terms

+

Other models

(GAM, Negative binominal model, etc.)

Important note
This soft was just made last week (1st version)

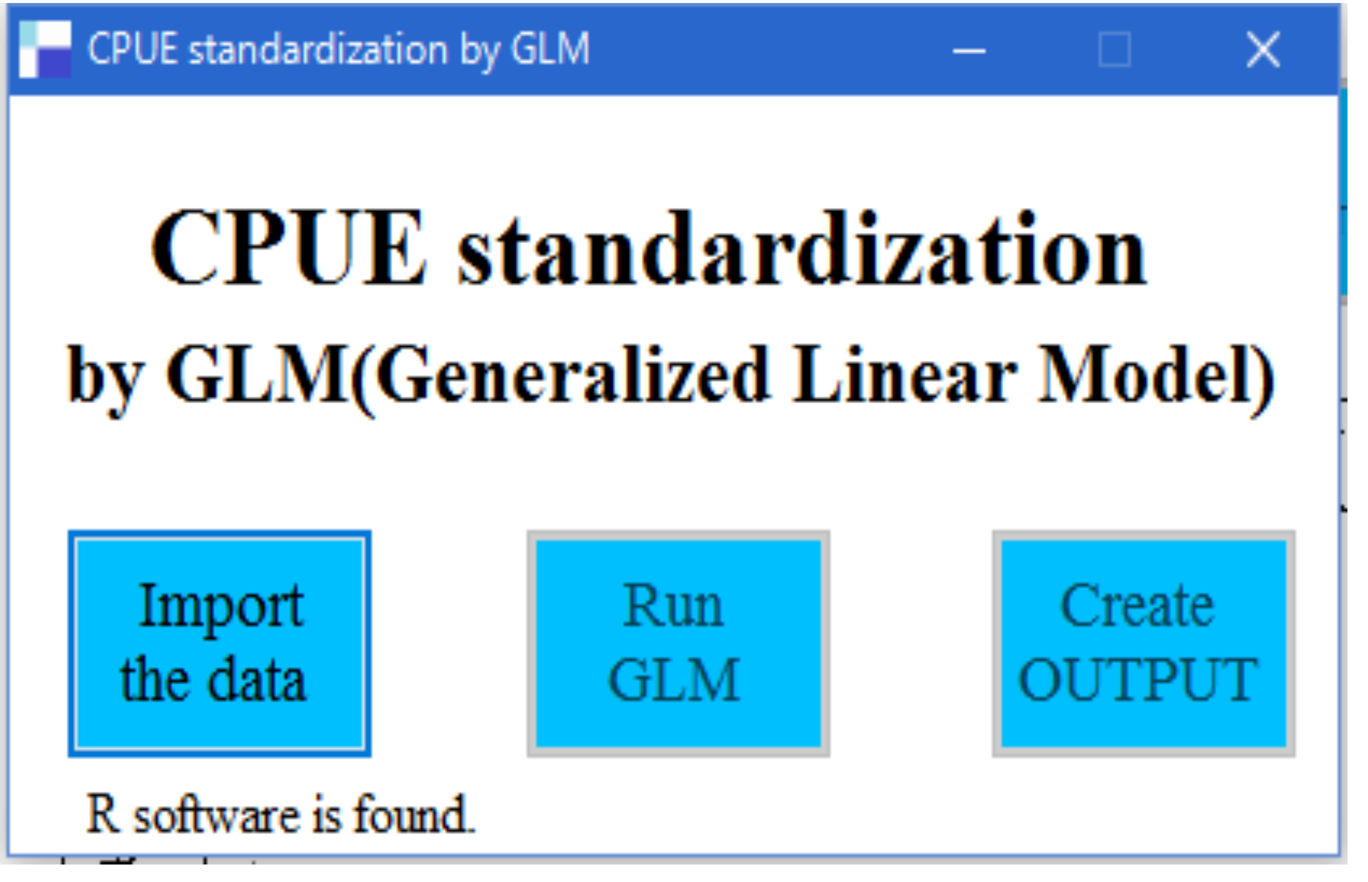
- We may need some improvements.
- We will note and later we will re-distribute the improved version after the training is completed.

How to set up Nominal CPUE data?

We need year, Season (Q or MO) and fishing area using Excel book file

	A	B	C	D
1	YR	Q	area	CPUE
2	2006	1	2	68.93996
3	2006	1	2	25.92267
4	2006	1	2	5.940594
5	2006	1	3	7.407407
6	2006	1	3	22.37979
7	2006	1	3	66.28212
8	2006	1	4	49.85229
9	2006	1	4	22.55489
10	2006	1	4	2.775675
11	2006	1	5	16.50444
12	2006	1	5	22.41932
13	2006	1	5	37.98127
14	2006	2	1	7.764953

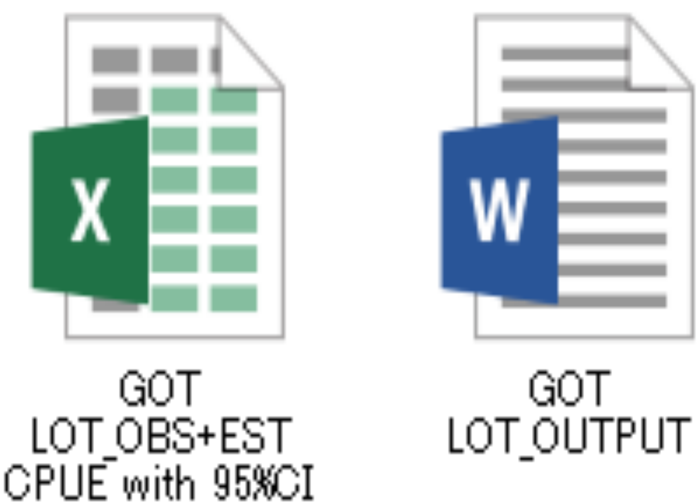
Main menu


CPUE standardization by GLM
CPUE standardization
by GLM(Generalized Linear Model)
Import
the data
Run
GLM
Create
OUTPUT
R software is found.

Try 3 sample data

How to interpret the outputs (most important)

- You will get 2 output files
- One Word file and One Excel file



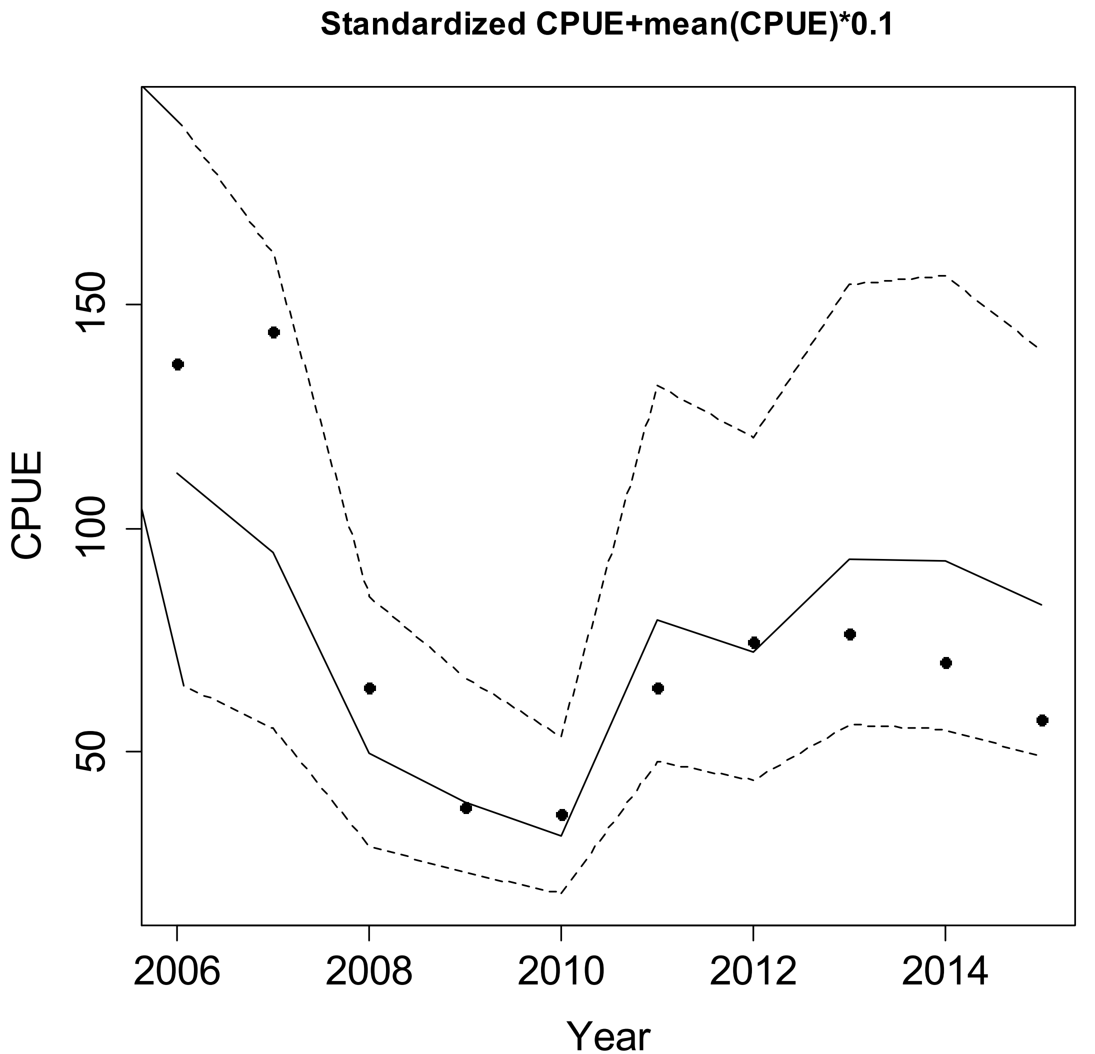
Word file: 4 types of output

[1] ANOVA table

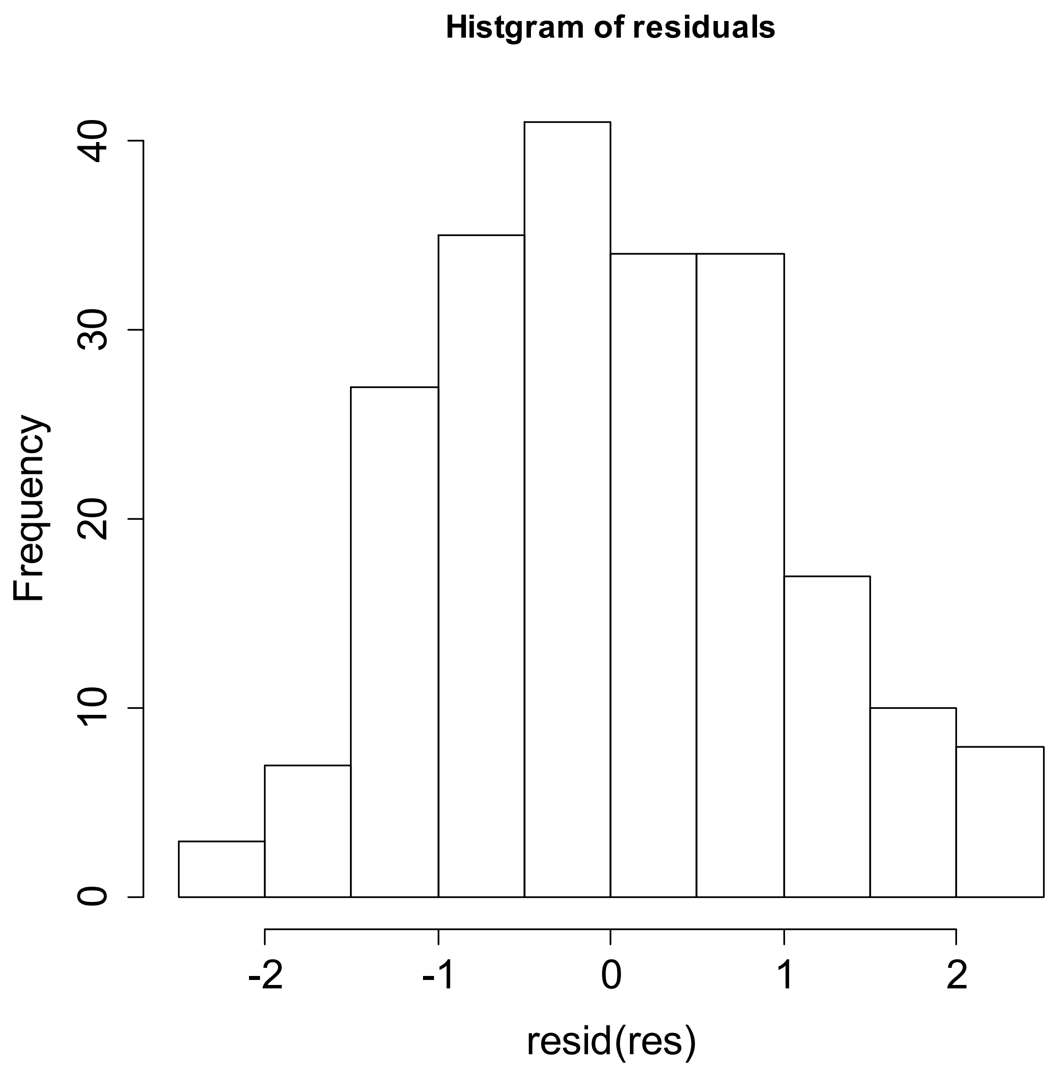
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Adjusted R2
as.factor (Data$YR)	9	33.758	3.7509	3.3986	7e-04	NA
as.factor (Data$Q)	3	48.2649	16.0883	14.5775	0	NA
as.factor (Data$area)	1	0.5977	0.5977	0.5415	0.4626	NA
Residuals	202	222.936	1.1036	NA	NA	0.2234

[2] Graph including 4 annual figures

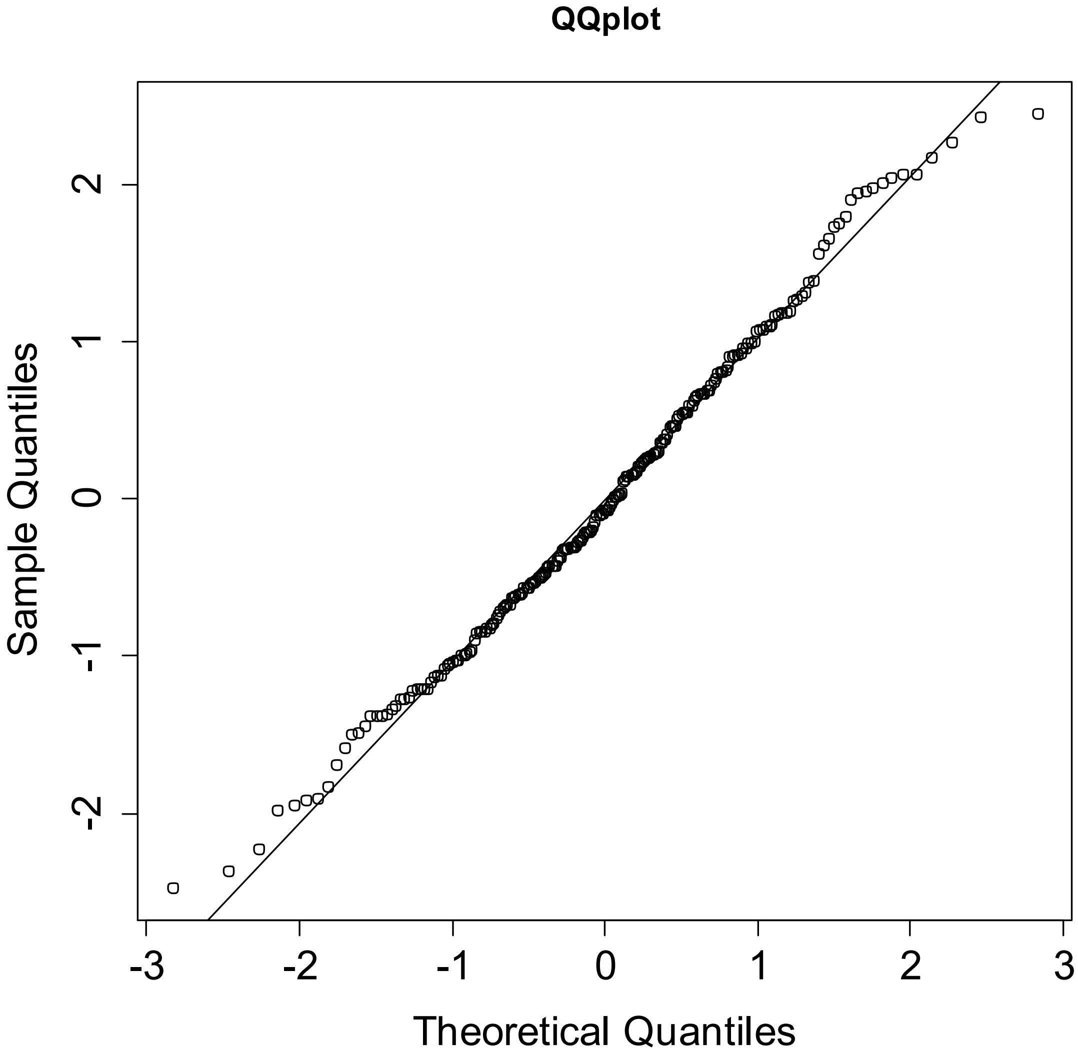
OBS (nominal) CPUE	Dot
EST (standardized) CPUE	Solid line
C: Lower limit of 95% CI	Dot line (Lower)
D: Upper limit of 95% CI.	Dot line (upper)



[2] Histogram of residuals



[3] QQ plot


QQplot
Sample Quantiles
Theoretical Quantiles
-2
-1
0
1
2
-3
-2
-1
0
1
2
3

How GLM is conducted by R?

```
#First, please set the working directory. This is the place of the file.
setwd("C:/")

#!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!#
#Loading data. Note that the order of column should be set as the same of the sample proram.
#In other words, prepare these column by this order, "Year", "season", "area" and "CPUE", although
#the name of the column is not the matter.
#!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!#

Data <- read.delim("GOT KAW.txt", header=T)
colnames(Data) <- c("YR", "Q", "area", "CPUE")

#Model fitting by linear model
#Here, we added the mean value of CPUE by 0.1 (i.e., mean(CPUE)*0.1) because there were a lot of zero values.
res <- lm(log(Data$CPUE+mean(Data$CPUE*0.1)) ~ as.factor(Data$YR) + as.factor(Data$Q) + as.factor(Data$area))
summary(res)

###Calculating estimated CPUE+mean(CPUE)*0.1 and its 95% confidential interval###
Data$Estimated_CPUE      <- exp(predict(res)+sd(summary(res)$residuals)^2/2)
Data$Estimated_CPUE975 <- exp(predict(res)+predict(res, se.fit=T)$se*1.96+sd(summary(res)$residuals)^2/2)
Data$Estimated_CPUE025 <- exp(predict(res)-predict(res, se.fit=T)$se*1.96+sd(summary(res)$residuals)^2/2)
```

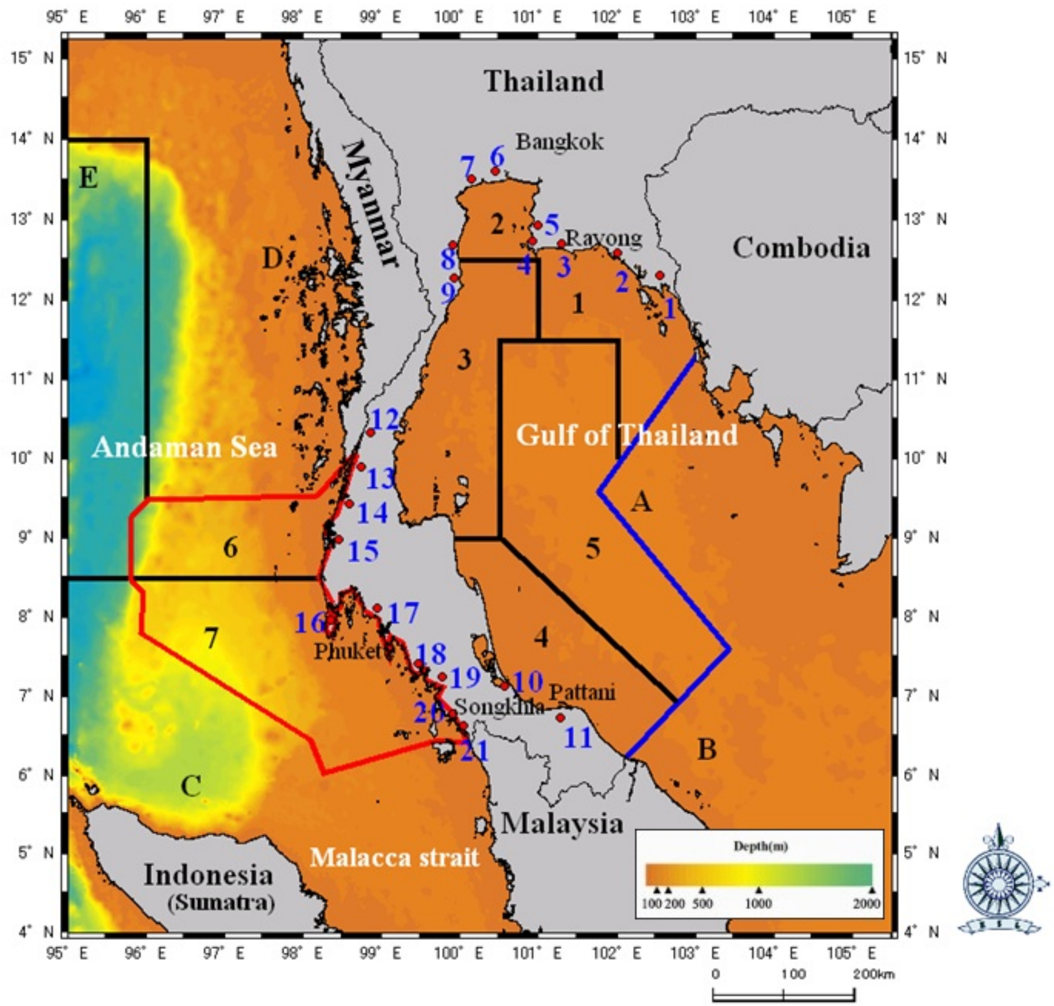
How GLM is conducted by R?

```
#If you want, you can check the estimated log(CPUE+mean(CPUE)*0.1) and its 95% confidential intervals
#plot(log(Data[, 4]+mean(Data[, 4])*0.1))
#points(predict(res), type="l")
#points(predict(res)+predict(res, se.fit=T)$se*1.96, type="l", lty=2)
#points(predict(res)-predict(res, se.fit=T)$se*1.96, type="l", lty=2)

#If you want, you can check the estimated CPUE+mean(CPUE)*0.1 and its 95% confidential intervals
#plot(Data[, 4]+mean(Data[, 4])*0.1, ylab="CPUE", cex.lab=1.5, cex.axis=1.5, pch=16)
#points(Data$Estimated_CPUE, type="l", lwd=1.2)
#points(Data$Estimated_CPUE025, type="l", lty=2, lwd=1.2)
#points(Data$Estimated_CPUE975, type="l", lty=2, lwd=1.2)

#Now, we estimated CPUE+mean(CPUE)*0.1 by year, area and season. However, we only need the CPUE by year.
#Therefore, the estimated CPUE was aggregated to yearly CPUE
Obs_CPUE <- aggregate(Data$CPUE+mean(Data$CPUE)*0.1~ Data$YR, FUN=mean,
na.rm=T)
Est_CPUE <- aggregate(Data$Estimated_CPUE~ Data$YR, FUN=mean, na.rm=T)
Est_CPUE025 <- aggregate(Data$Estimated_CPUE025~ Data$YR, FUN=mean, na.rm=T)
Est_CPUE975 <- aggregate(Data$Estimated_CPUE975~ Data$YR, FUN=mean, na.rm=T)

#Estimated yearly CPUE+mean(CPUE)*0.1 and its 95% confidential interval.
#Bold line is estimated yearly CPUE+mean(CPUE)*0.1 and black points are observed yearly
CPUE+mean(CPUE)*0.1.
#Break lines show the 95% confidential interval.
setwd("C:/") #the place of output file
pdf("Outputs GOT KAW.pdf") #output as PDF file
plot(Obs_CPUE[, 1], Obs_CPUE[, 2], ylim=c(min(Obs_CPUE, Est_CPUE025[, 2]), max(Obs_CPUE[, 2],
Est_CPUE975[, 2])),
    xlab="Year", ylab="CPUE", cex.lab=1.5, cex.axis=1.5, pch=16, main="Standardized CPUE+mean(CPUE)*0.1")
points(Est_CPUE[, 1], Est_CPUE[, 2], type="l", lwd=1.5)
points(Est_CPUE025[, 1], Est_CPUE025[, 2], type="l", lty=2, lwd=1.5)
points(Est_CPUE975[, 1], Est_CPUE975[, 2], type="l", lty=2, lwd=1.5)
```

```
#Histgram of residuals (diferrences between estimated and observed CPUE+mean(CPUE)*0.1)
hist(resid(res), cex.lab=1.5, cex.axis=1.5,main="Histgram of residuals")

#QQplot
qqnorm(resid(res),   cex.lab=1.5, cex.axis=1.5, main="QQplot")
qqline(resid(res))
dev.off()

#Anova table
resanova <- anova(res, test="F")

#R2
r2 <- summary(res)$adj.r.squared

M <- matrix(NA, nrow=nrow(resanova), ncol=(ncol(resanova)+1))
M[, 1] <- round(resanova[, 1], 4)
M[, 2] <- round(resanova[, 2], 4)
M[, 3] <- round(resanova[, 3], 4)
M[, 4] <- round(resanova[, 4], 4)
M[, 5] <- round(resanova[, 5], 4)
M[nrow(M), 6] <- round(r2, 4)
rownames(M) <- rownames(resanova)
colnames(M) <- c(colnames(resanova), "Adjusted R2")

#Anova table and adjusted R2 are output as CSV file
write.csv(M, "ANOVA and r2 GOT KAW.csv")

#The standardized and nominal CPUE
M2 <- matrix(NA, nrow=nrow(Obs_CPUE), ncol=4)
rownames(M2) <- Obs_CPUE[, 1]
M2[, 1] <- Obs_CPUE[, 2]
M2[, 2] <- Est_CPUE[, 2]
M2[, 3] <- Est_CPUE025[, 2]
M2[, 4] <- Est_CPUE975[, 2]
colnames(M2) <- c("Observed CPUE+0.1*mean(CPUE)", "Estimated CPUE+0.1*mean(CPUE)", "2.5%", "97.5%")
write.csv(M2, "Standardized CPUE0.1mean(CPUE) GOT KAW.csv")
```

Practice data process to create nominal CPUE

Thailand CPUE
Gulf of Thailand and Andaman Sea

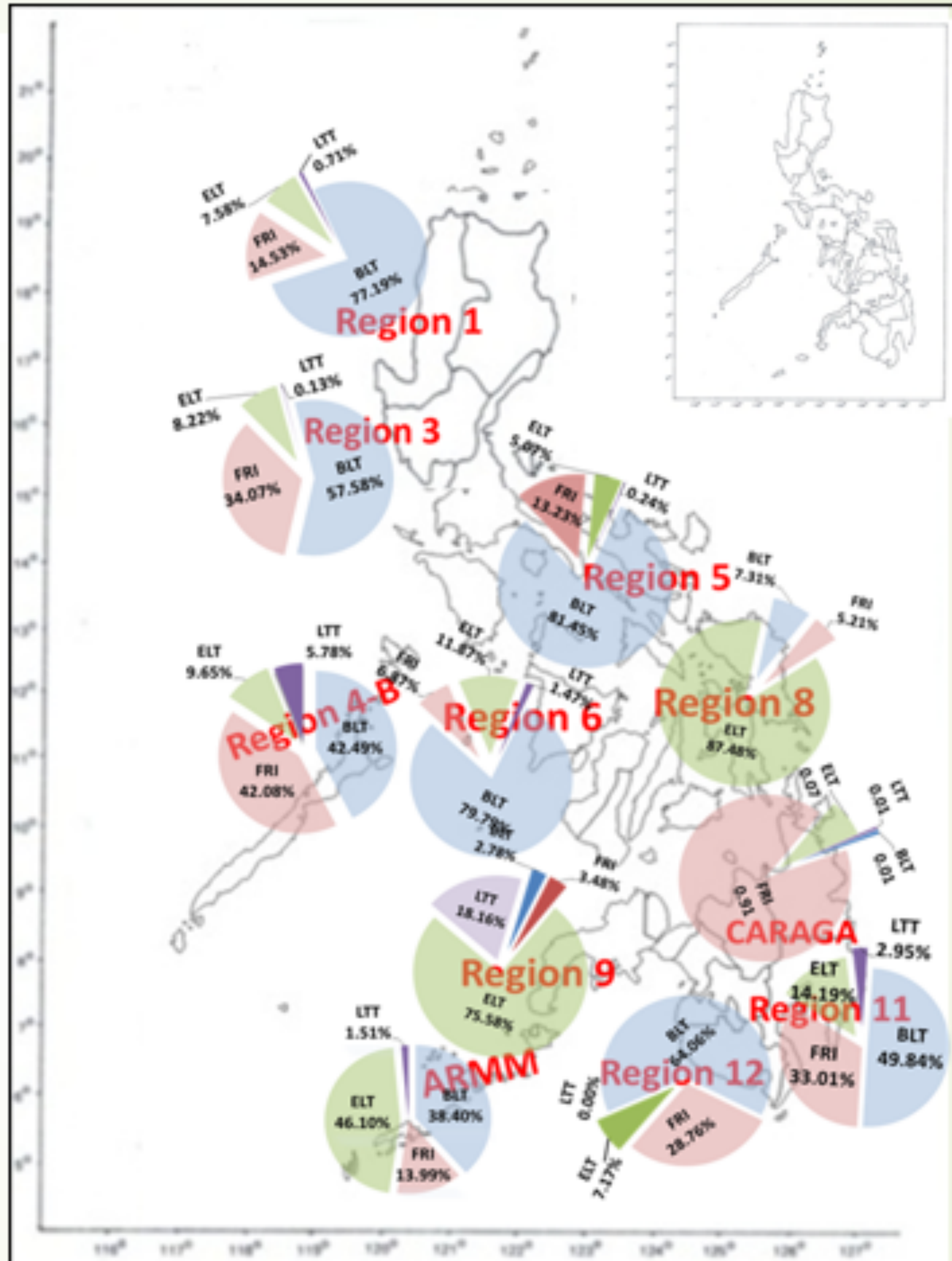


Philippines CPUE

Spp. Dominance:

- **longtail tuna** – Region 4B, 9, 11
- **eastern little tuna** - Regions 8, 9 and ARMM.

Data received Region1-6 and 11 (not major area for LOT and KAW) → 0 catch



Philippines C+E data Received (Region 1-5+11)

Data are not from major fishing grounds

(LOT:4B+9+11)

(LOT: even major area ➔ very few catch)

(bycatch) (not possible to apply by GLM)

(KAW: 8+9+ARMM)

bycatch area Many 0 catches

(not possible to apply by GLM)

If you apply GLM for CE data with many 0 catches

We will have very unstable nominal CPUE

As well as standardized CPUE

Not realistic

Can not be used for ASPIC

and any other stock assessment models

There are some techniques for bycatch (many 0 catch) CPUE standardization (beyond this training course)

% of 0 (zero) catch	Model
less than 20%	GLM
20-60%	negative binominal
60% or more	0 inflated 2 set negative binominal

There are many other references (future considerations)

- Prof. Shono (Supopon’s teacher)

Reference (2004) in Japanese

A Review of Some Statistical Approaches Used for CPUE Standardization

Hiroshi SHONO

Bull. Jpn. Soc. Fish. Oceanogr.

Anyway

- We practice Thai and Philippines catch and effort data