

# まぐろはえ縄データを用いた 標準化CPUEの推定



独立行政法人

水産総合研究センター

遠洋水産研究所 くらまぐろ資源部

太平洋くらまぐろ資源研究室

市野川 桃子

# 日本のはえ縄漁業のCPUE

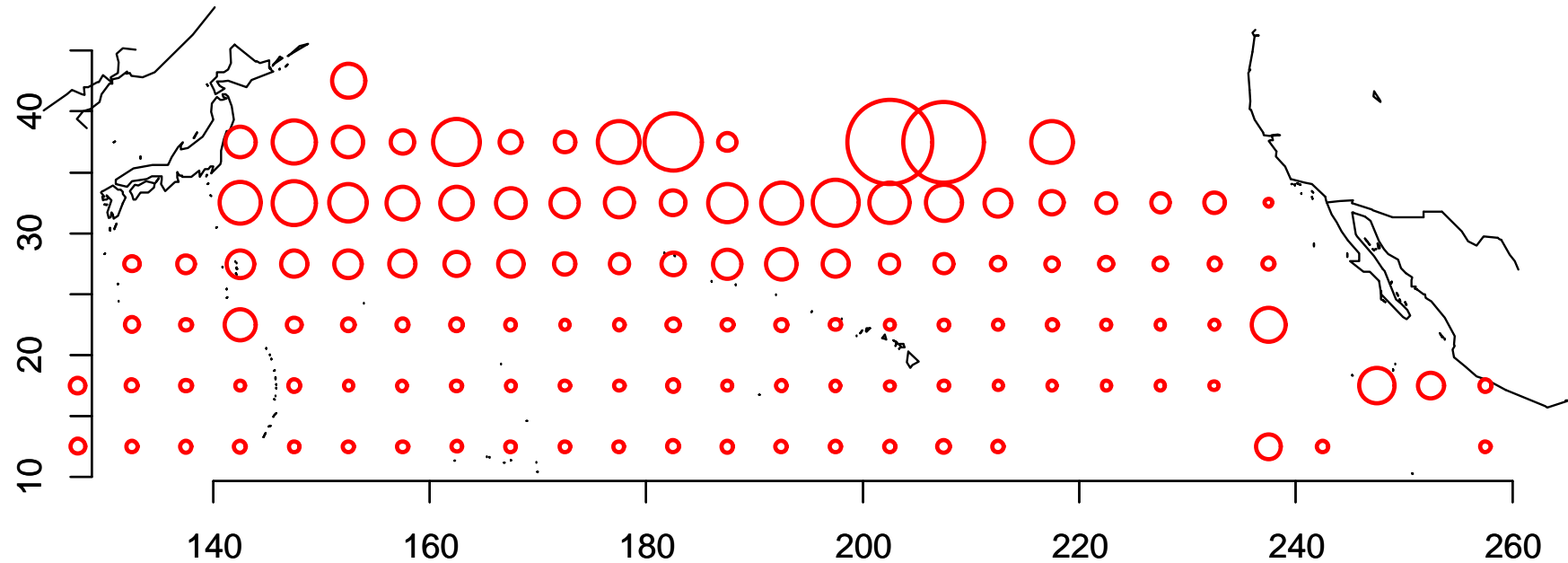
- 諸外国に比べて精度が良い
- 漁場のカバー率が広い
- 50年以上の長期間にわたって整備

多種の資源評価で、  
最も重要なインプット  
データとして利用

漁場や季節、年代によって  
対象種の釣れやすさが  
異なる可能性がある  
→ その効果が、資源量指数を  
バイアスさせるのを防ぐ必要  
(CPUEの標準化)

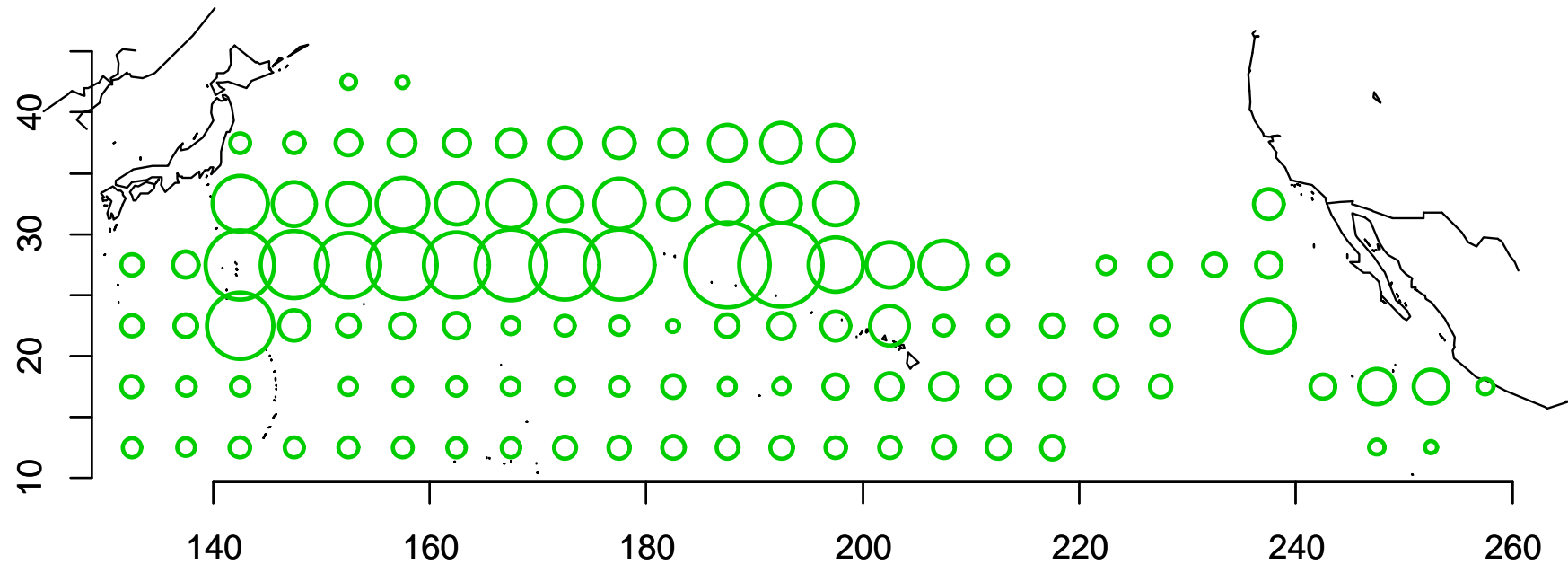
# はえ縄漁業データの例

メカジキのノミナルCPUEの分布 1~3月



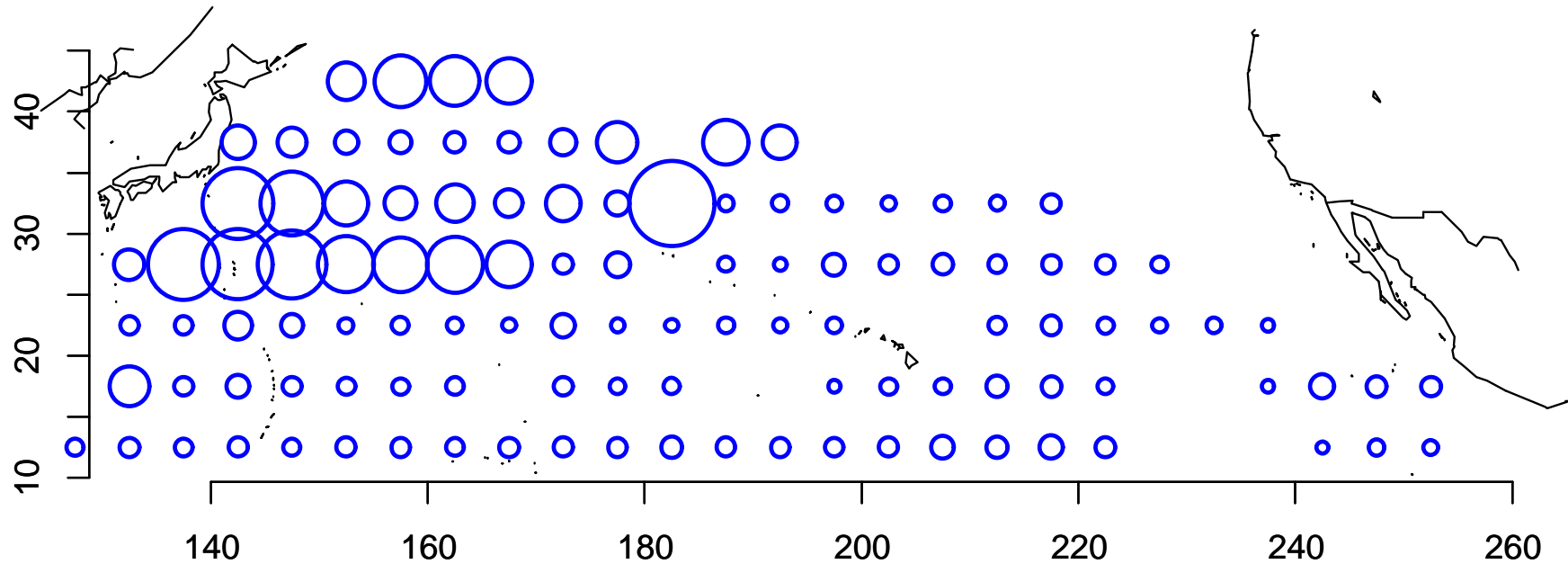
# はえ縄漁業データの例

メカジキのノミナルCPUEの分布 4~6月



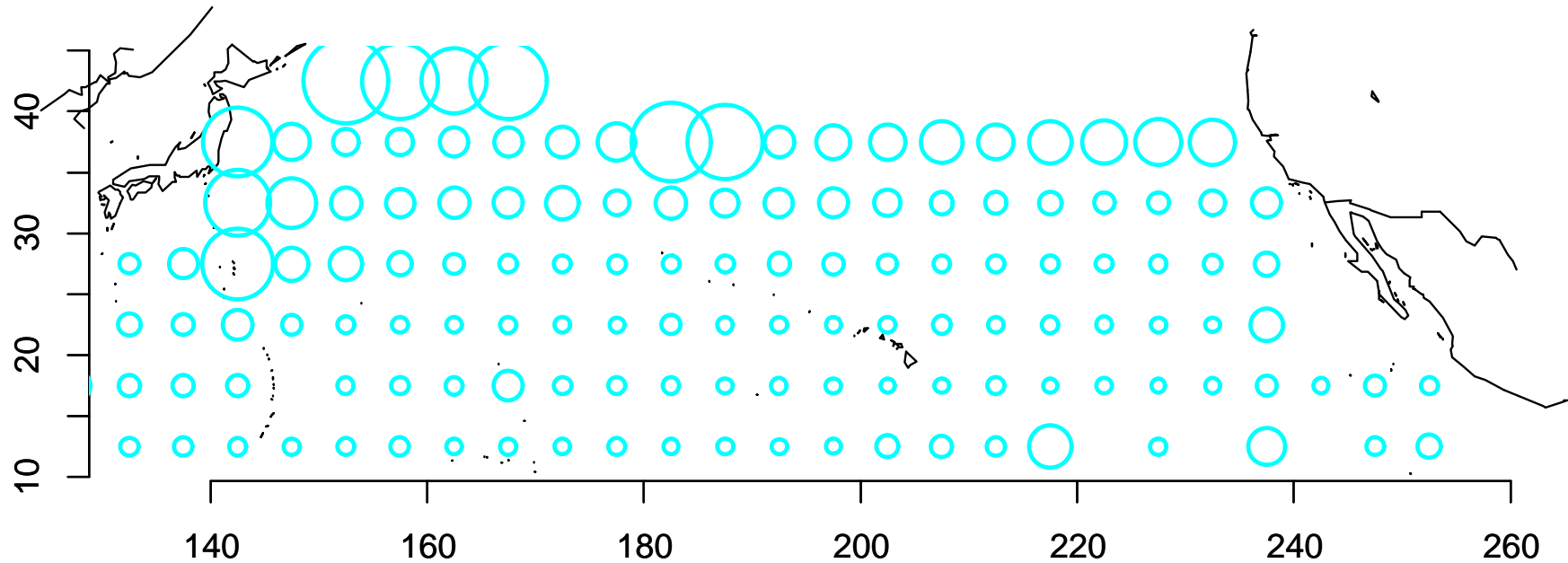
# はえ縄漁業データの例

メカジキのノミナルCPUEの分布 7~9月



# はえ縄漁業データの例

メカジキのノミナルCPUEの分布 9~12月



# 本発表の概要

はえ縄漁業データでよく行われる

「CPUEの標準化」の紹介

- ① 標準化CPUEとは？（要旨の内容）
  - 標準化CPUEとは？
  - 標準化CPUEの推定
  - 推定以前・以後の注意点
  
- ② サンプルデータを使った実際の手順の紹介（要旨にはない内容）
  - SASとRを使った場合
  - 標準化の効果の検討

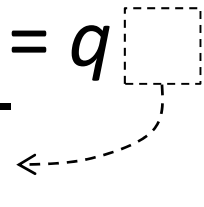
# ① 標準化CPUEとは？



# CPUEとは？

## 資源動向を示す指標

ある年( $y$ )の漁獲量( $C_y$ )は、  
〔 海の資源( $N_y$ )が多いほど、  
〔 漁獲のために費やす努力量 ( $E_y$ ) が多いほど、  
多くなる

$$C_y = q E_y N_y$$
$$CPUE_{E_y} = \frac{C_y}{E_y} = q \boxed{\phantom{N_y}} N_y$$


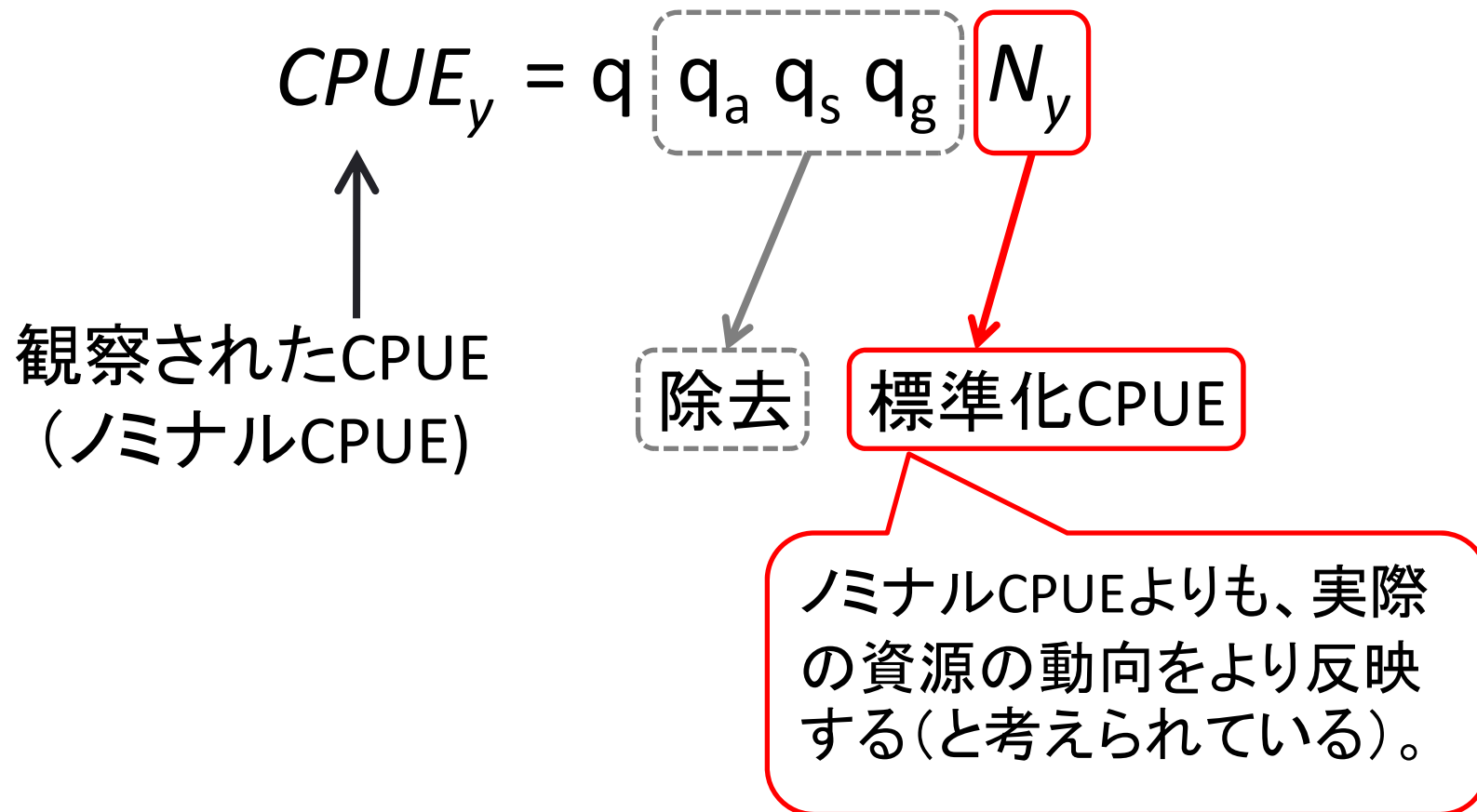
→ CPUE ( $C_y/E_y$ ) は、資源量と比例関係にある

漁獲量は、努力量と資源量だけで決まる？  
海域・漁具・季節によって魚の獲れ具合は違う

ある年( $y$ )の漁獲量( $C_y$ )は、  
海の資源( $N_y$ )が多いほど、  
漁獲のために費やす努力量 ( $E_y$ ) が多いほど、  
良い漁場 ( $q_a$ ) や季節 ( $q_s$ )、漁具 ( $q_g$ ) を  
選べば選ぶほど、  
多くなる

$$CPUE_y = \frac{C_y}{E_y} = q \quad q_a \quad q_s \quad q_g \quad N_y$$

# 余計な効果の除去 = CPUEの標準化



## GLMを用いた標準化CPUEの推定

$$CPUE_y = q \cdot q_a \cdot q_s \cdot q_g \cdot N_y$$

対数をとる

$$\log(CPUE_y) = \log(q) + \log(q_a) + \log(q_s) + \log(q_g) + \log(N_y) + \text{誤差項}$$

観察されたCPUE  
(ノミナルCPUE)

観察データを説明する  
線形モデル (GLM)

観察されたCPUEと予測値の差(残差)が仮定した誤差分布と似るようにパラメータ( $q, q_a, \dots$ )を推定

# 統計ソフトによるGLMの実行

## ○ SASの場合

```
proc glm ;  
  class year area;  
  model LCPUE = year area / ss1 ss3 solution;  
  lsmeans year area / stderr out=estim;  
run;
```

## ○ Rの場合

```
res <- glm(logcpue~effect1 + .. + as.factor(year),  
           data=fishery.data)
```

- GLMの実行は、準備がきちんとしていれば、コマンドを1行打つだけ！
- GLMの推定結果が妥当であれば、GLMの結果から資源量指数としての標準化CPUEを取り出すことができる

## glmのコマンド以前・以後 (1)

- glm (またはlm)のコマンドを打つのは簡単  
→ 真の資源量は不明なので、結果の妥当性の十分な検討が必要

統計学的・漁業学的・生物学的妥当性

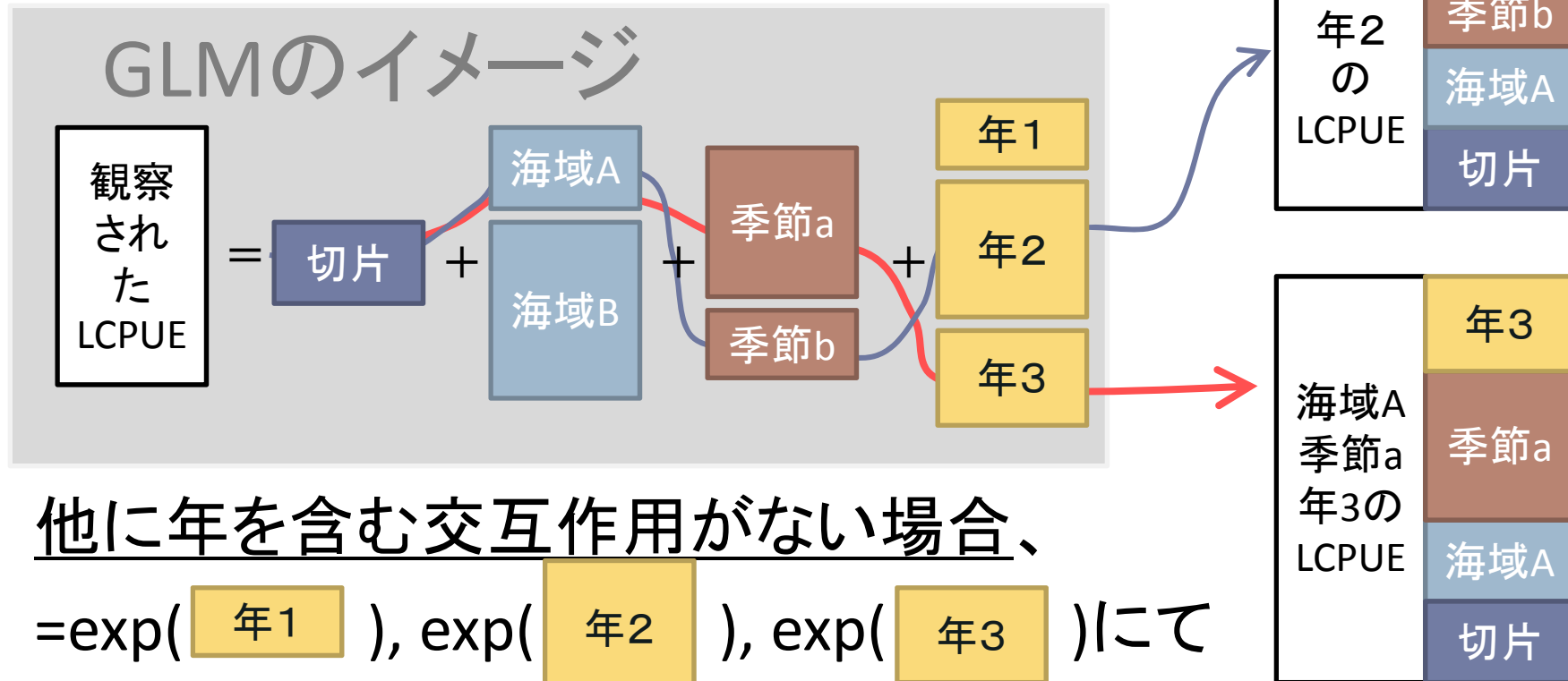
- 情報量基準を用いたモデル選択
- 残差の偏りの確認
- ゼロキャッチデータの適切な処理

- 重要と思われる効果(漁具)が正しく導入されているか？
- 年代によって漁獲効率が変わるか？
- 適切な努力量を用いているか？(→ 金岩先生の発表)

- 実際の生息域や回遊パターンを考慮、海域や季節の説明変数を導入できているか？

# 年トレンドの抽出①

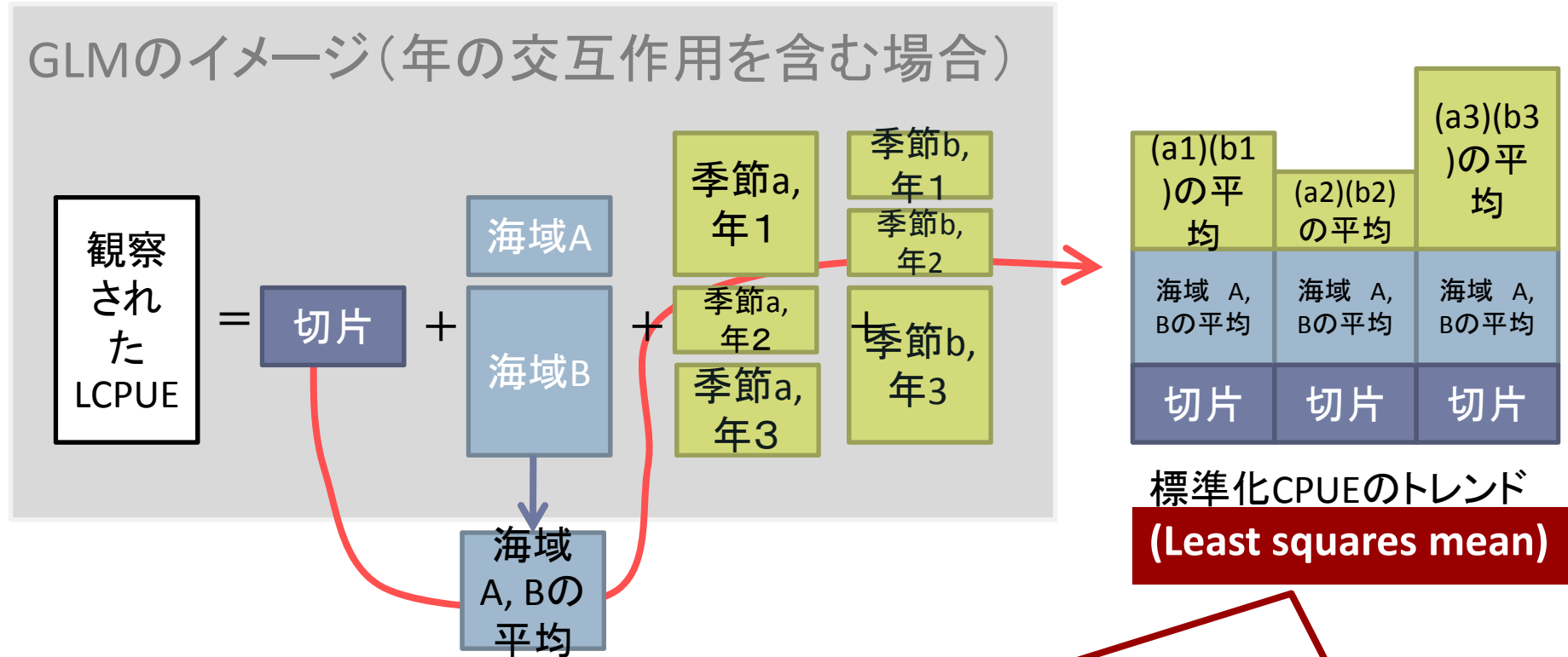
得られたパラメータの何が「標準化CPUE」にあたるのか？



他に年を含む交互作用がない場合、

=exp( 年1 ), exp( 年2 ), exp( 年3 )にて  
年のトレンドは抽出できる

## 年トレンドの抽出②



Least squares mean (LS mean) の推定値とその標準誤差は、SASでは簡単に計算できるが、Rでは難しいのが悩み。

→ サンプルデータで、計算例を紹介します

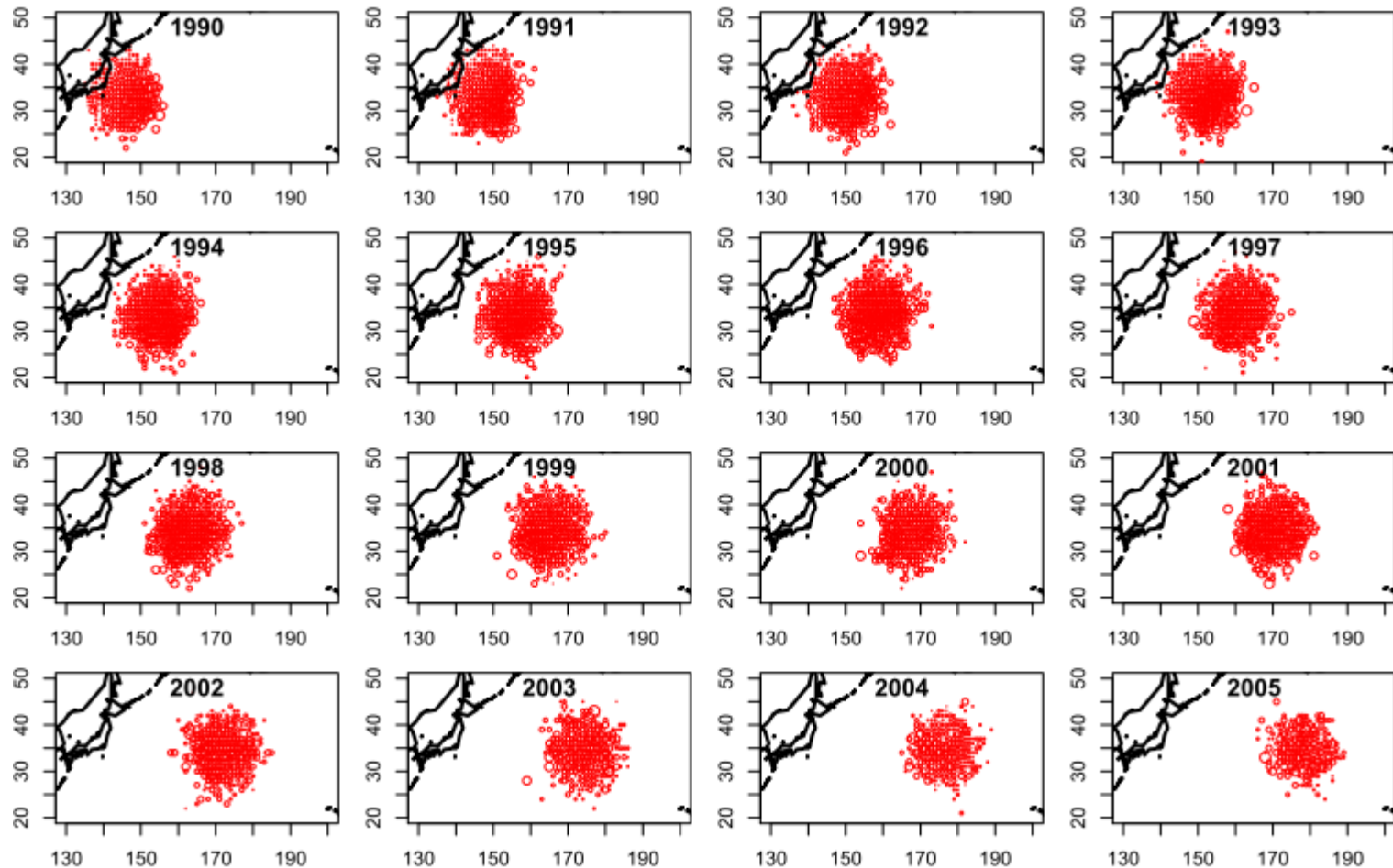


## ② サンプルデータを使った 解析例

# サンプルデータを使った解析例

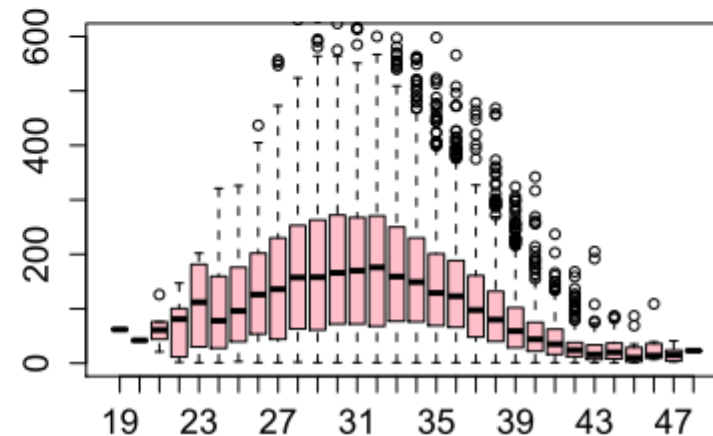
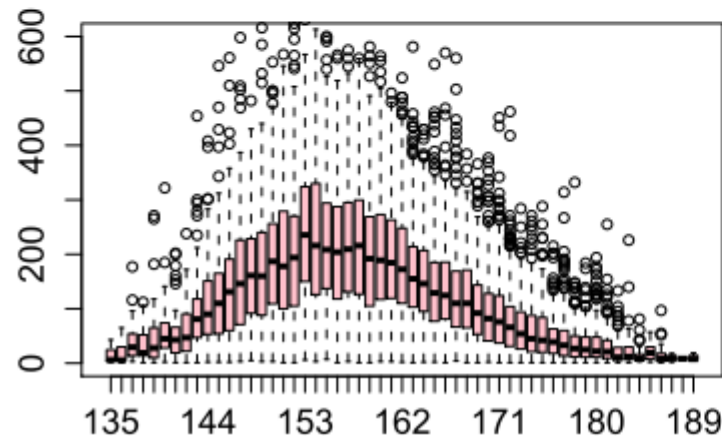
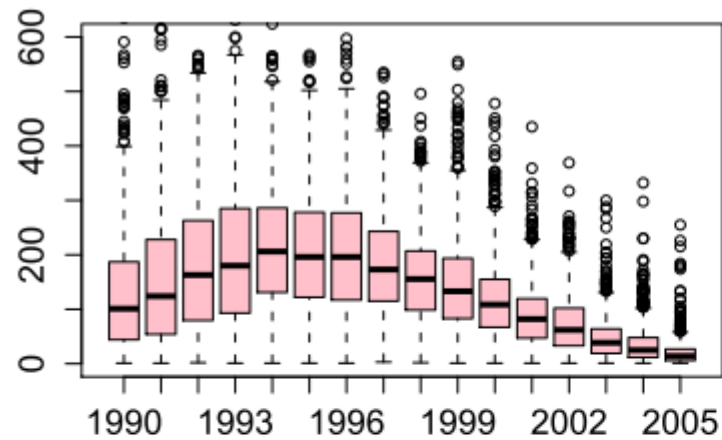
- データ; testdata (1990年から2005年)
- CPUE・年・緯度・経度のデータ (シミュレーションにより作成)

CPUEの分布



# サンプルデータを使った解析例

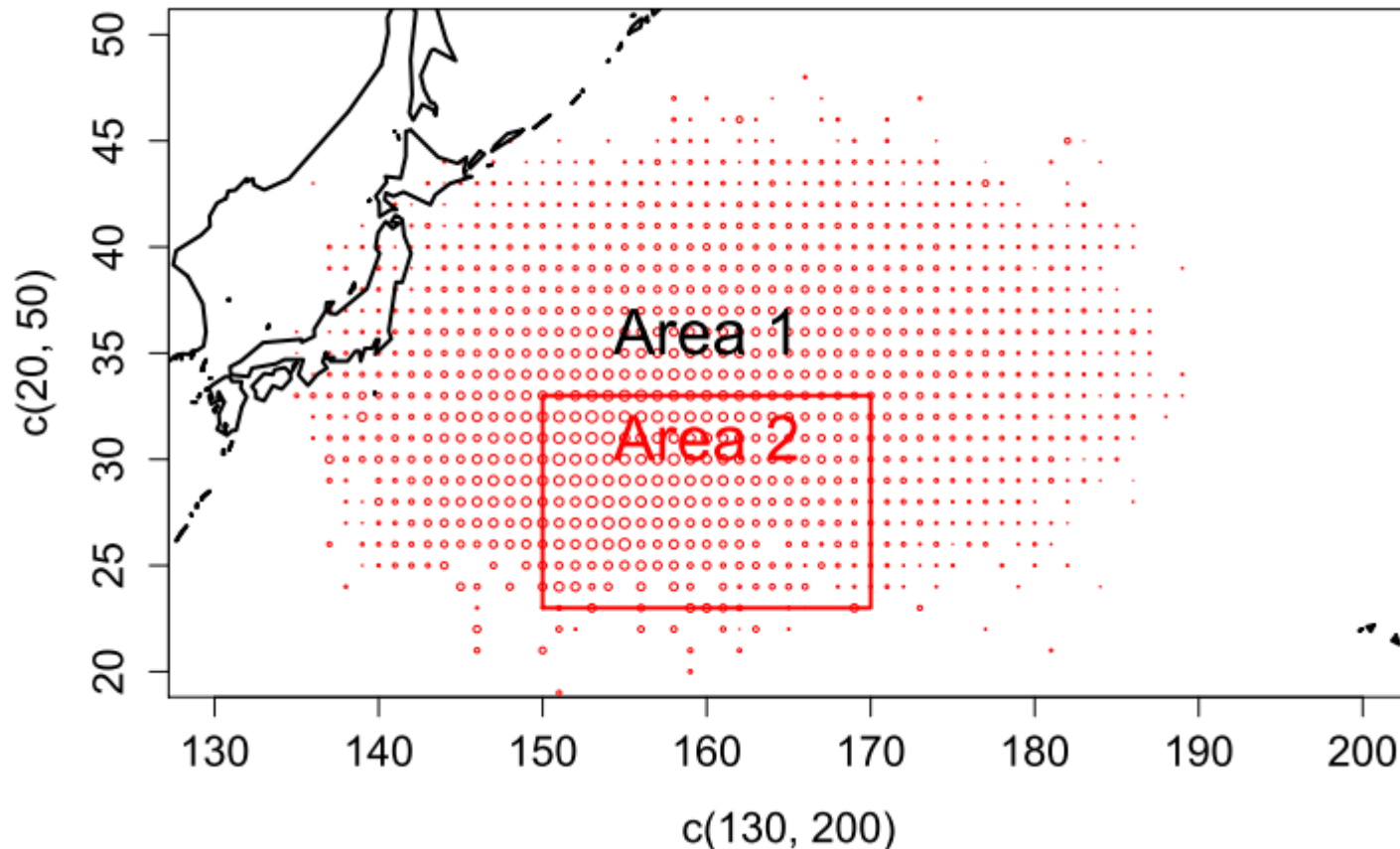
- データ; testdata (1990年から2005年)
- CPUE・年・緯度・経度のデータ(シミュレーションにより作成)



# サンプルデータを使った解析例

- 仮定するモデル

$$\text{LCPUE} (\log (\text{CPUE})) = \text{year (カテゴリカル)} + \text{area (カテゴリカル)} \\ \sim \text{正規分布の誤差}$$



## ② サンプルデータを使った解析例 (SAS)

### SASプログラム

```
proc glm ;  
  class year area;  
  model LCPUE = year area / ss1 ss3 solution;  
  lsmeans year area / stderr out=estim;  
run;
```

### SAS出力例

The GLM Procedure

		Class Level Information											
Class	Levels	Values											
year	16	1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001											
		2002 2003 2004 2005											
area	2	1 2											
		Number of Observations											
		Number of Observations											

基本的情報の表示

Source	DF	Squares	Mean Square	F Value	Pr > F
Model	16	8604.77899	537.79869	696.19	<.0001
Error	15221	11757.96456	0.77248		
Corrected Total	15237	20362.74356			

R-Square Coeff Var Root MSE LCPUE Mean  
0.422575 19.66069 0.878910 4.470394

Source	DF	Type I SS	Mean Square	F Value	Pr > F
year	15	7625.809743			
area	1	978.969248			

分散分析表 (ss1, ss3)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	15	5707.397381			
area	1	978.969248			

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	3.197492110 B	0.03792725	84.31	<.0001
year 1990	1.829973787 B	0.02998239	61.05	<.0001
year 1991	1.937481322 B	0.02998239	64.66	<.0001
year 1992	2.217617087 B	0.02998239	73.97	<.0001
..... 省略 .....				
year 1999	1.959524624 B	0.02998239	65.36	<.0001
year 2000	1.797995353 B	0.02998239	59.99	<.0001
year 2001	1.587107555 B	0.04378569	36.25	<.0001
year 2002	1.325896615 B	0.04380755	30.27	<.0001
year 2003	0.870503765 B	0.04418454	19.70	<.0001
year 2004	0.519431701 B	0.04481145	11.59	<.0001
year 2005	0.000000000 B			
area 1	-0.639706549 B	0.01796970	-35.60	<.0001
area 2	0.000000000 B			

推定されたパラメータ

The GLM Procedure  
Least Squares Means  
Standard

year	LCPUE	LSMEAN	Error	Pr >  t
1990	4.70761262	0.02998239	<.0001	
1991	4.81512016	0.02998239	<.0001	
1992	5.09525592	0.02998239	<.0001	
1993	5.13134271	0.02998239	<.0001	
1994	5.22555355	0.02998239	<.0001	
1995	5.15113030	0.02998239	<.0001	
1996	5.16579278	0.02998239	<.0001	
1997	5.08736521	0.02998239	<.0001	
1998	4.96625912	0.02799280	<.0001	
1999	4.83716346	0.02795722	<.0001	

LS mean (標準化CPUE) とその標準誤差

## ② サンプルデータを使った解析例 (SAS→Excel)

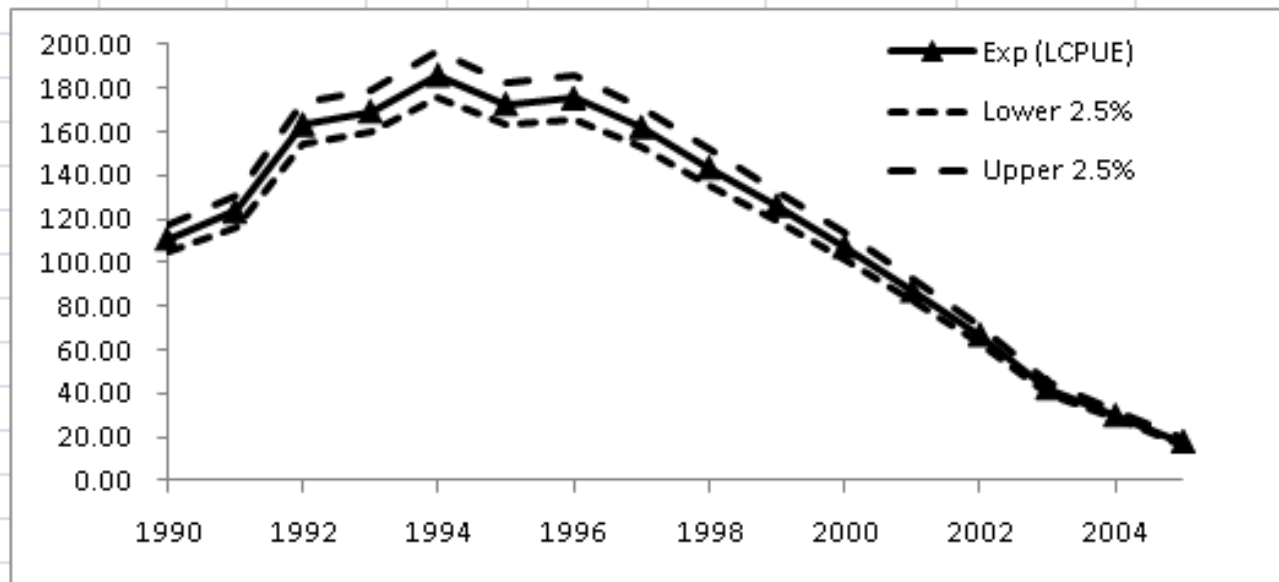
### Excel 計算例 (資源量指数のプロット)

year	LCPUE LSMEA	Error	Pr> t	Lower 2.5%	Upper 2.5%	Exp (LCPUE)	Lower 2.5%	Upper 2.5%
1990	4.708	0.0300	<.0001	4.648	4.768	110.84	104.339	117.634
1991	4.815	0.0292	<.0001	4.757	4.874	123.41	116.358	130.787
1992	5.095	0.0287	<.0001					
1993	5.131	0.0283	<.0001					
2005	2.878	0.0346	<.0001					

SAS出力結果  
からコピー

エクセルの関数で計算

- ・ ログスケールから普通スケールに  
(  $\text{Exp (LCPUE)} = \exp(\text{LCPUE LSMEAN}) + (\text{Error}^2)/2$  )
- ・ 95% 信頼区間の計算  
 $95\% \text{ conf} = \exp(\text{LCPUE LSMEAN} \pm 2 * \text{Error})$



## ② サンプルデータを使った解析例 (Rの場合)

### ～説明変数とデータ数の確認～

#### SAS出力例

The GLM Procedure

Class	Levels	Values	Class
year	16	1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005	
area	2	1 2	

Source		152
Model		152
Error		152
Corrected Total		152

R-Square  
0.422575

#### Rプログラム&出力

```
> res <- lm(logcpue~as.factor(year) + as.factor(area),  
data=testdata)  
  
> lmres1$xlevels  
$`as.factor(year)`  
[1] "1990" "1991" "1992" "1993" "1994" "1995" "1996" "1997"  
"1998" "1999"  
[11] "2000" "2001" "2002" "2003" "2004" "2005"  
$`as.factor(area)`  
[1] "1" "2"  
  
> nrow(testdata)  
[1] 15238  
  
> summary(lmres1)$adj.r.squared  
[1] 0.4219677
```

## ② サンプルデータを使った解析例 (R with SASの結果)

### ～分散分析表 (Type 1, Sequential)～

#### SAS出力例

Source	DF	Type I SS	Mean Square	F Value	Pr > F
year	15	7625.809743	508.387316	658.12	<.0001
area	1	978.969248	978.969248	1267.30	<.0001

#### Rプログラム&出力

```
> anova(lmres1,test="F")
Analysis of Variance Table
Response: LCPUE
          Df Sum Sq Mean Sq F value  Pr(>F)
as.factor(year)  15 7625.8  508.39 658.12 < 2.2e-16 ***
as.factor(area)   1  979.0  978.97 1267.30 < 2.2e-16 ***
Residuals    15221 11758.0   0.77
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



## ② サンプルデータを使った解析例 (R with SASの結果)

### ～分散分析表 (Type III, adjusted SS)～

#### SAS出力例

Source	DF	Type III SS	Mean Square	F Value	Pr > F
year	15	5707.397381	380.493159	492.56	<.0001
area	1	978.969248	978.969248	1267.30	<.0001

#### Rプログラム&出力

```
> library(car) # library car のインストール (install.packages("car") が必要)
```

```
> Anova(lmres1,type="III",test.statistic="F")
```

```
Anova Table (Type III tests)
```

```
Response: LCPUE
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	5490.4	1	7107.49	< 2.2e-16 ***
as.factor(year)	5707.4	15	492.56	< 2.2e-16 ***
as.factor(area)	979.0	1	1267.30	< 2.2e-16 ***
Residuals	11758.0	15221		

```
---
```

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

## ② サンプルデータを使った解析例 (R with SASの結果)

### ～推定された係数～

#### SAS出力例

```
Standard
Parameter      Estimate
Intercept      3.197492110 B
year 1990      1.829973787 B
year 1991      1.937481322 B
..... 省略 .....
year 2000      1.797995353 B
year 2001      1.587107555 B
year 2002      1.325896615 B
year 2003      0.870503765 B
year 2004      0.519431701 B
year 2005      0.000000000 B
area 1         -0.639706549 B
area 2         0.000000000 B
```

#### Rプログラム&出力

```
> summary(lmres1)$coefficients
```

	Estimate	Std. Error	t value
(Intercept)	4.38775935	0.02884648	152.1072448
as.factor(year)1991	0.10750754	0.04044298	2.6582493
as.factor(year)1992	0.38764330	0.04036648	9.6030980
..... 省略 .....			
as.factor(year)2001	-0.24286623	0.04029682	-6.0269335
as.factor(year)2002	-0.50407717	0.04034060	-12.4955303
as.factor(year)2003	-0.95947002	0.04076952	-23.5340029
as.factor(year)2004	-1.31054209	0.04145709	-31.6120161
as.factor(year)2005	-1.82997379	0.04419352	-41.4081888
as.factor(area)2	0.63970655	0.01796970	35.5991841

## ② サンプルデータを使った解析例 (R with SASの結果)

### ～推定された係数 (SAS風の出力)～

#### SAS出力例

```
Standard
Parameter      Estimate
Intercept      3.197492110 B 0
year 1990      1.829973787 B 0
year 1991      1.937481322 B 0
..... 省略 .....
year 2000      1.797995353 B 0
year 2001      1.587107555 B 0
year 2002      1.325896615 B 0
year 2003      0.870503765 B 0
year 2004      0.519431701 B 0
year 2005      0.000000000 B 0
area 1         -0.639706549 B 0.0
area 2         0.000000000 B .
```

#### Rプログラム&出力

```
> options("contrasts"=c("contr.SAS","contr.SAS"))
> lmres1.sas <- glm(LCPUE~(as.factor(year)+as.factor(area)
                    data=testdata)
> summary(lmres1.sas)$coefficient
              Estimate Std. Error    t value
(Intercept)    3.1974921 0.03792725  84.30594
as.factor(year)1990 1.8299738 0.04419352  41.40819
as.factor(year)1991 1.9374813 0.04389721  44.13678
..... 省略 .....
as.factor(year)2000 1.7979954 0.04392934  40.92926
as.factor(year)2001 1.5871076 0.04378569  36.24717
as.factor(year)2002 1.3258966 0.04380755  30.26639
as.factor(year)2003 0.8705038 0.04418454  19.70155
as.factor(year)2004 0.5194317 0.04481145  11.59150
as.factor(area)1    -0.6397065 0.01796970 -35.59918
```

## ② サンプルデータを使った解析例 (R with SASの結果)

### ～LS meanの計算～

#### SAS出力例

```
The GLM Procedure
      Least Squares Means
              Standard
year  LCPUE LSMEAN      Error  Pr > |t|
1990  4.70761262  0.02998239  <.0001
1991  4.81512016  0.02922636  <.0001
1992  5.09525592  0.02869906  <.0001
1993  5.13112016  0.02869906  <.0001
1994  5.22612016  0.02869906  <.0001
1995  5.15112016  0.02869906  <.0001
1996  5.16612016  0.02869906  <.0001
1997  5.08712016  0.02869906  <.0001
1998  4.96612016  0.02869906  <.0001
1999  4.83712016  0.02869906  <.0001
2000  4.67612016  0.02869906  <.0001
2001  4.46512016  0.02869906  <.0001
2002  4.20412016  0.02869906  <.0001
2003  3.74812016  0.02869906  <.0001
2004  3.39712016  0.02869906  <.0001
2005  2.87812016  0.02869906  <.0001
```

#### Rプログラム&出力

```
> tmp <- table(testdata$year,testdata$area) # 説明変数として用いた各層のデータ数
> dummy.data <- as.data.frame.table(x) # データフレームに変換
> colnames(dummy.data) <- c("year","area","Freq") # 対応する名前をつける
# 全層に努力量が均一であるという擬似データから得られるモデルの予測値を平均
> lsmean.year.area <-
      tapply(predict(lmres1,newdata=dummy.data,list(x$year,x$area),mean)
> lsmean.year <- apply(lsmean.year.area,1,mean)
> round(lsmean.year,3)
1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002
4.708 4.815 5.095 5.131 5.226 5.151 5.166 5.087 4.966 4.837 4.676 4.465 4.204
2003 2004 2005
3.748 3.397 2.878
```

## ② サンプルデータを使った解析例 (R with SASの結果)

### ～LS meanの標準誤差の計算(ブートストラップ)～

#### Rプログラム&出力

```
> N.boot <- 5000
> lsmean.boot <- matrix(0,length(unique(year)),N.boot)
> n.obs <- nrow(bootdata)
> for(i in 1:N.boot){
  bootdata <- resample.strata(testdata,year)
  tmpres <- lm(LCPUE~as.factor(year)-1,data=bootdata)
  lsmean.boot[i,] <-
    apply(tapply(predict(tmpres,newdata=bootdata[,year]),year,
    FUN=function(x){lm(LCPUE~as.factor(x)-1,data=bootdata[,year])$residuals}),
    MARGIN=2,SE=TRUE)
}
> lsmean.sd <- sqrt(apply(lsmean.boot,1,VAR))
> round(lsmean.sd,3)
[1] 0.03516 0.03541 0.02693 0.02545
[10] 0.02369 0.02478 0.02497 0.0301
```

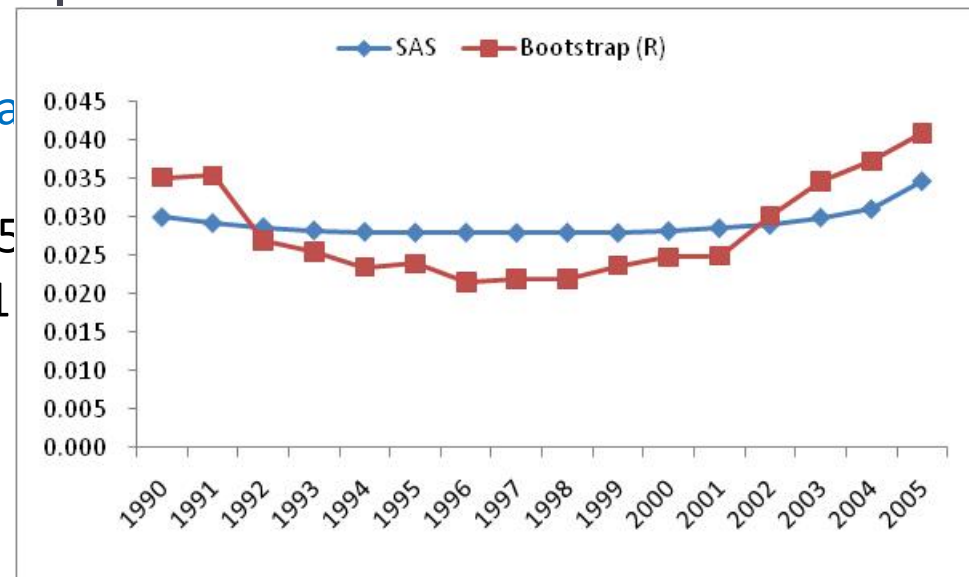
#### SAS出力例

The GLM Procedure

Least Squares Means

Standard

year	LCPUE	LSMEAN	Error	Pr >  t
1990	4.70761262	0.02998239	<.0001	
1991	4.81512016	0.02922636	<.0001	
1992	5.09525592	0.02869906	<.0001	
1993	5.13134271	0.02827566	<.0001	
1994	5.22555355	0.02807519	<.0001	
1995	5.15113030	0.02798747	<.0001	



## ② サンプルデータを使った解析例 (R with SASの結果)

### Rプログラム (層別リサンプリングの関数)

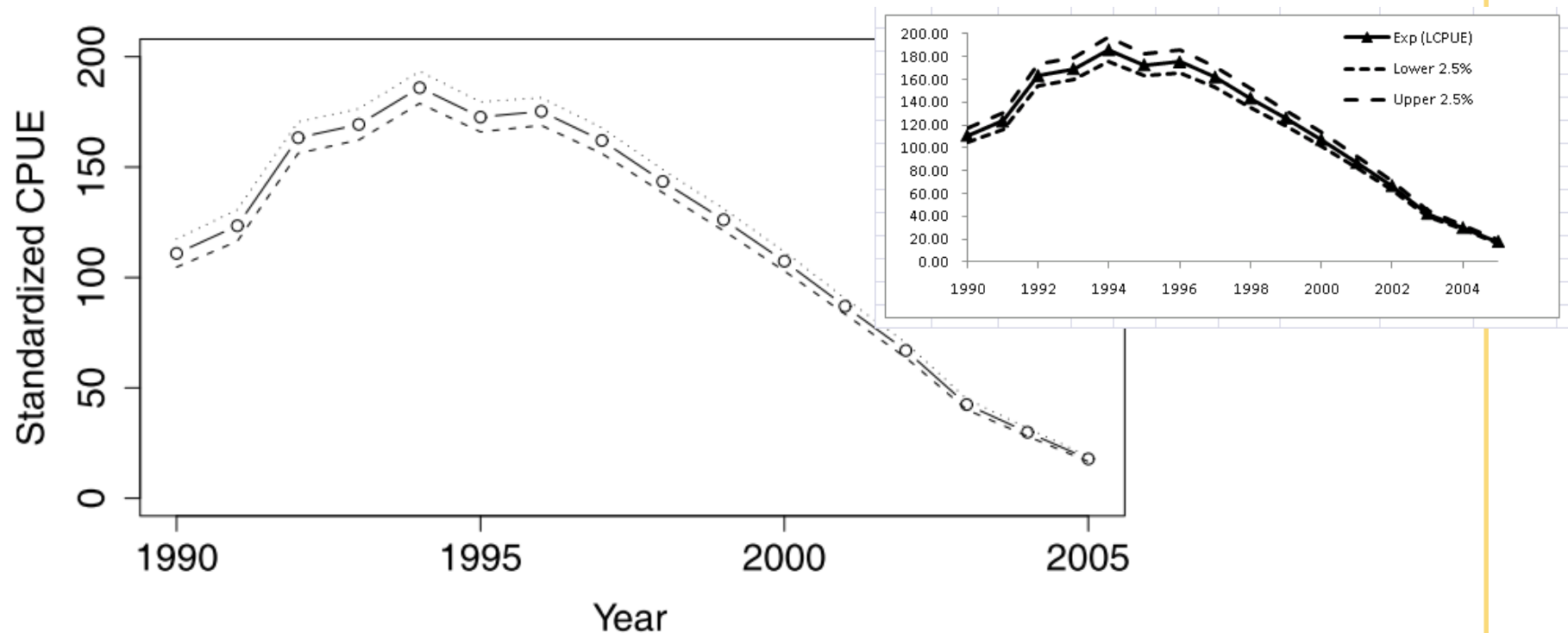
```
resample.strata <- function(odata,nsample=1){
  yr.tmp <- sort(unique(odata$year))
  area.tmp <- sort(unique(odata$area))
  rdata <- rep(0,ncol(odata))
  for(i in 1:length(yr.tmp)){
    for(j in 1:length(area.tmp)){
      tmp <- odata$year==yr.tmp[i] & odata$area==area.tmp[j] #&
odata$gear==gear.tmp[k]
      data.tmp <- odata[tmp,]
      if(sum(tmp)>0){
        rdata <- rbind(rdata,
data.tmp[sample(1:nrow(data.tmp),ceiling(nrow(data.tmp)*nsample),replace=
T),])
      }
    }
  }
  rdata <- rdata[-1,]; colnames(rdata) <- colnames(odata)
  return(rdata)
}
```

## ② サンプルデータを使った解析例 (R with SASの結果)

### ～標準化CPUEのプロット～

#### Rプログラム&出力

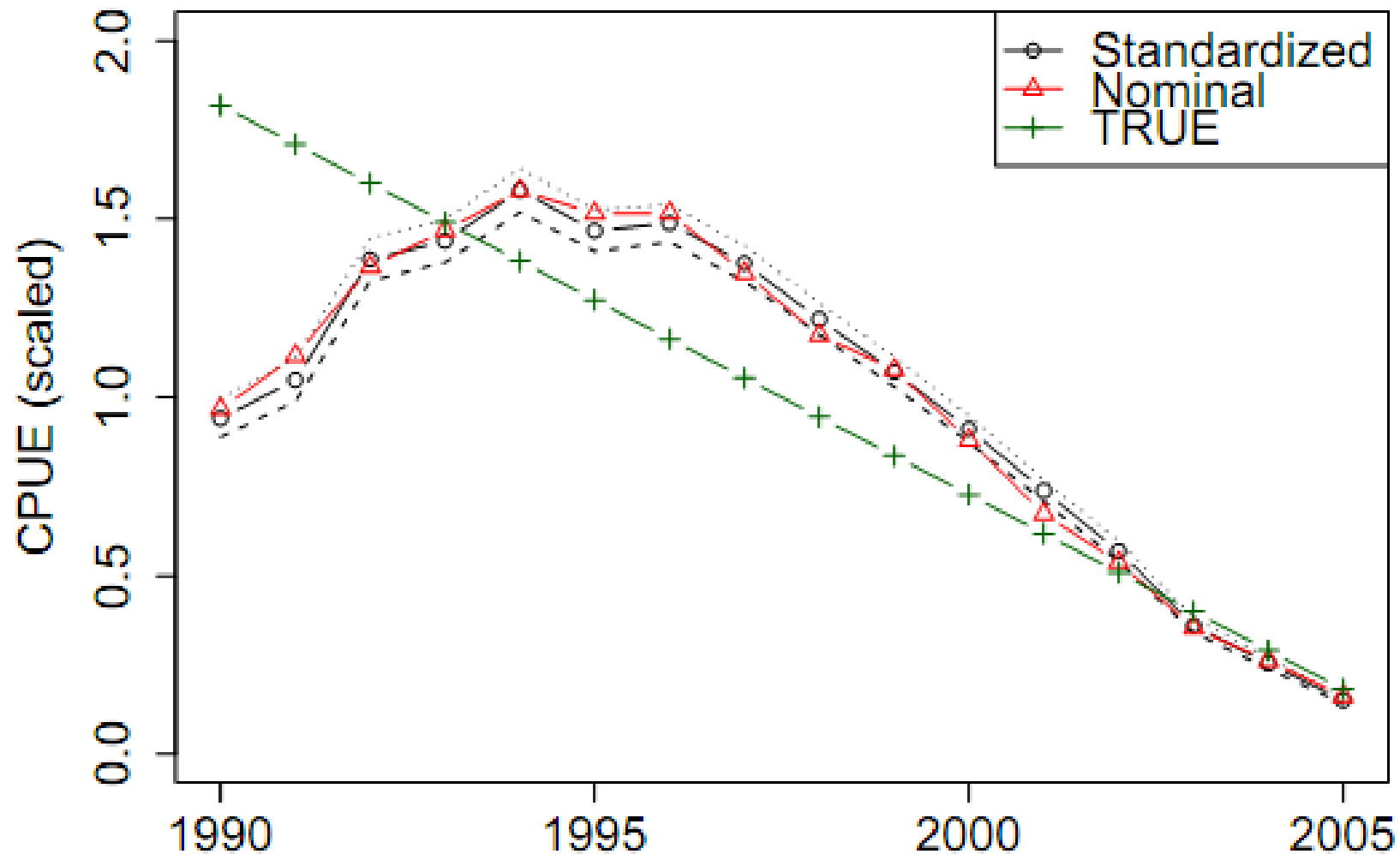
```
> plot(names(lsmmean.year),  
       exp(lsmmean.year+(lsmmean.sd^2)/2),type="b")  
> lines(names(lsmmean.year),  
        exp(apply(lsmmean.boot,1,quantile,probs=0.05))),type="l",col=1)  
> lines(names(lsmmean.year),  
        exp(apply(lsmmean.boot,1,quantile,probs=0.95))),type="l",col=1)
```



## ② サンプルデータを使った解析例 (結果)

(答えあわせ)

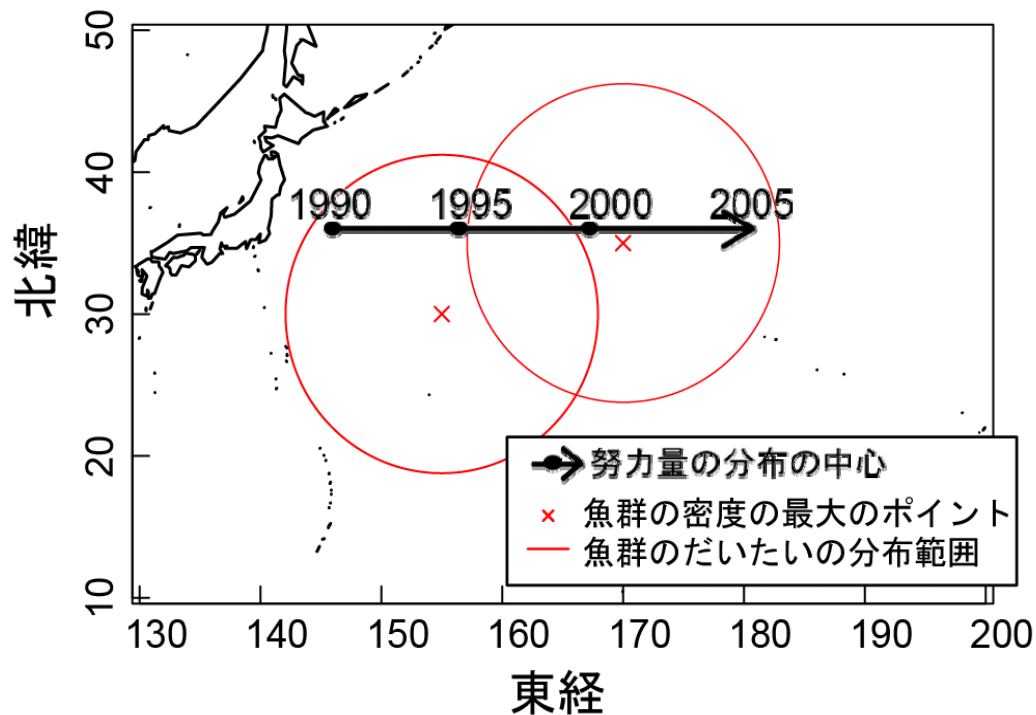
推定された標準化CPUEと真の資源量





# シミュレーションモデルのシナリオ

- ・ 漁業は1990年からスタートし、日本沿岸からだんだん沖に漁場がシフト

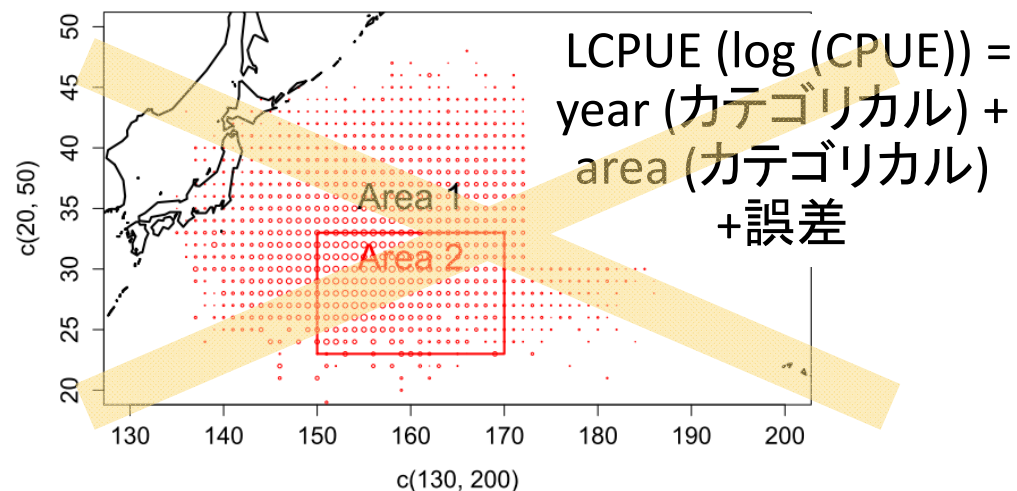


## 漁業データ

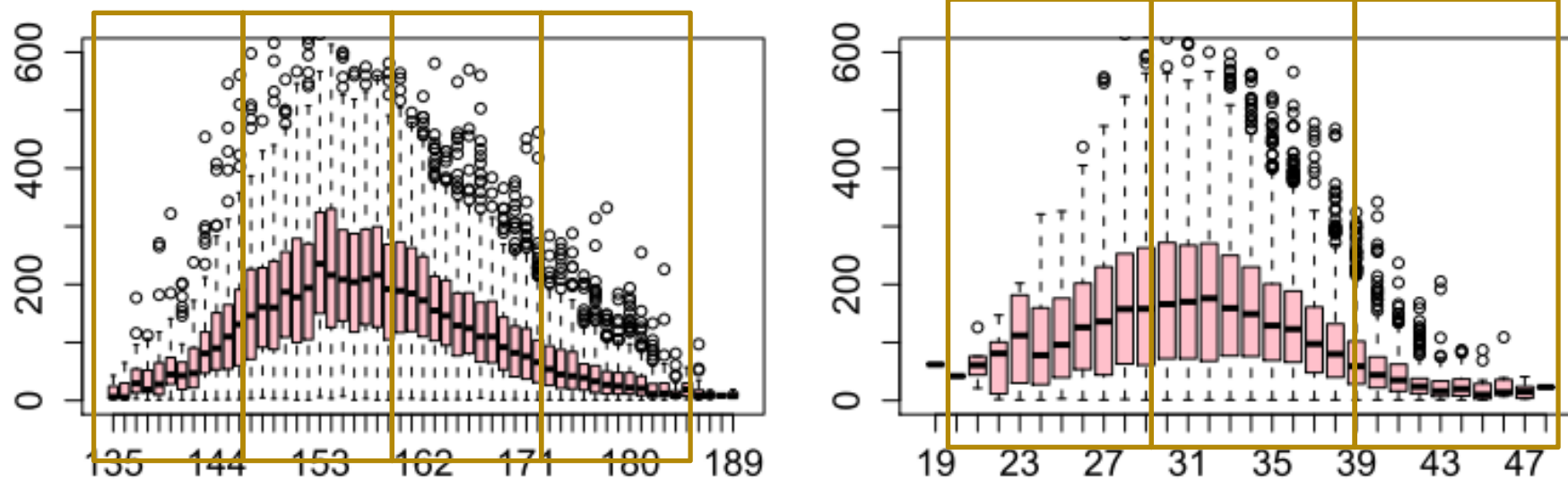
- ・ 操業位置 (緯度・経度)
- ・ 操業年
- ・ 操業あたりの漁獲尾数
- ・ 一年で 1000 操業
- ・ testdata.csv という形で保存

## ② サンプルデータを使った解析例 (結果)

何がいけなかったのか？

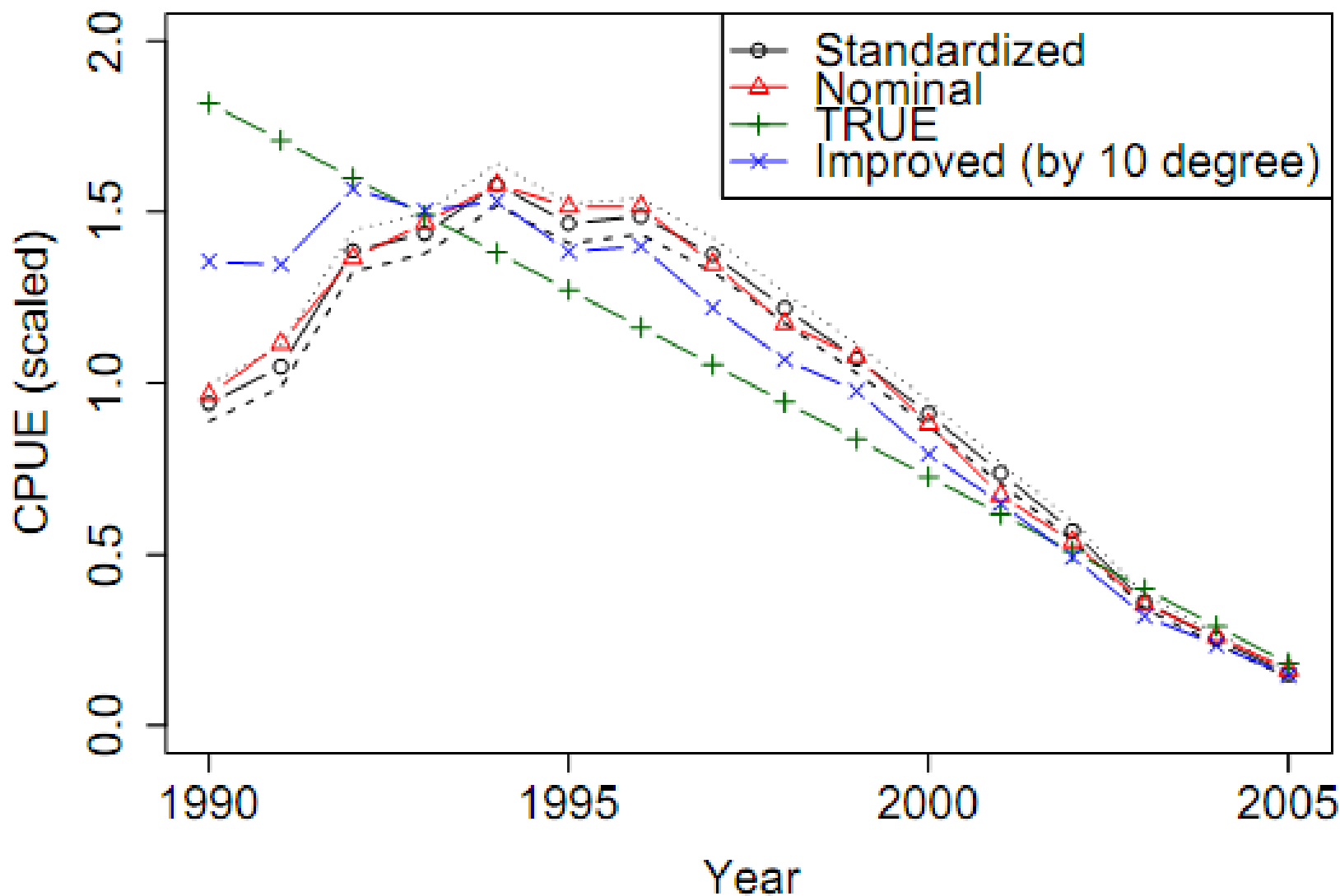


改善モデル案 (緯度・経度をカテゴリカル変数として導入)

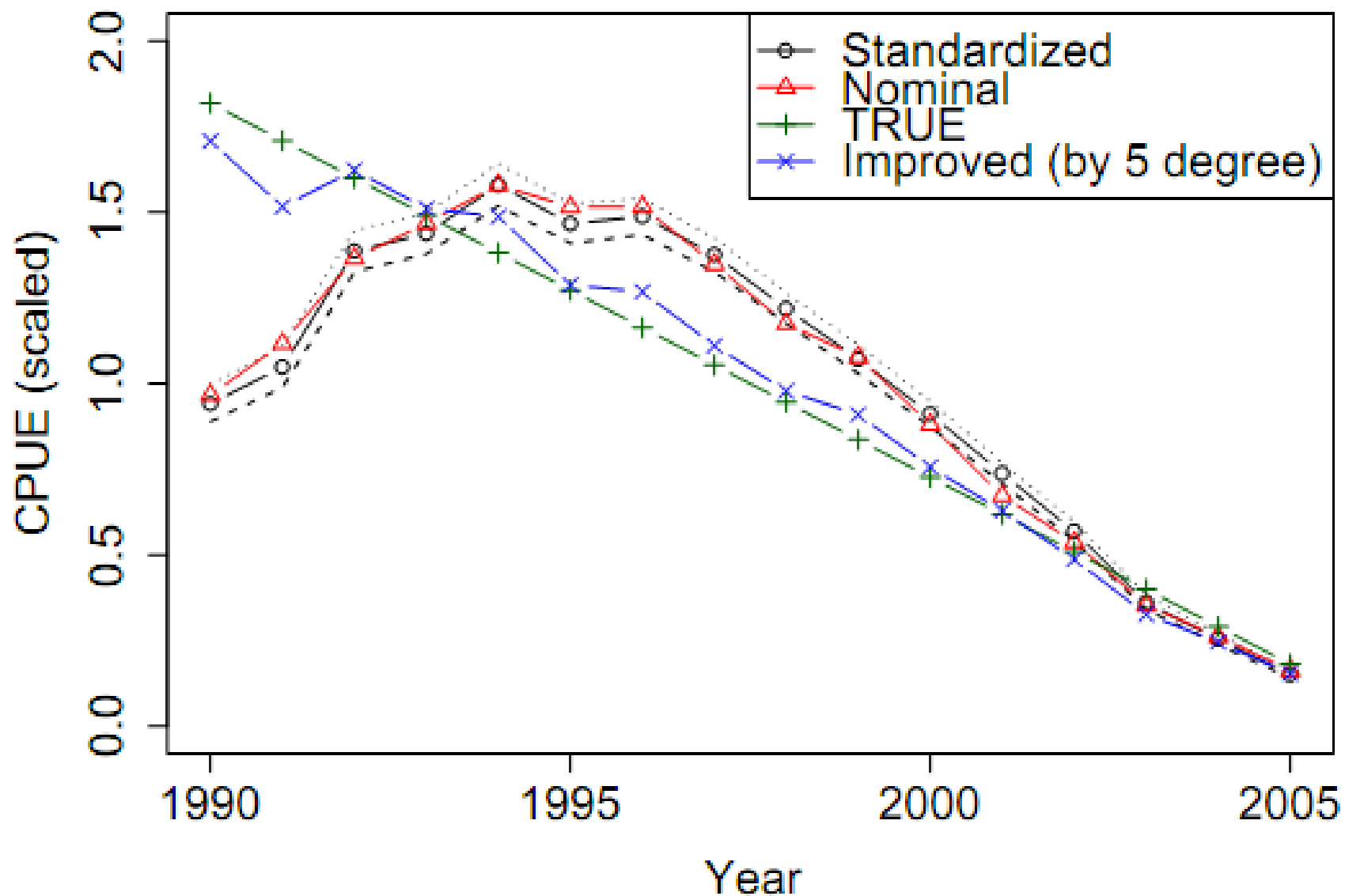


LCPUE (log (CPUE)) = year (カテゴリカル) + lon (カテゴリカル) + lat (カテゴリカル)

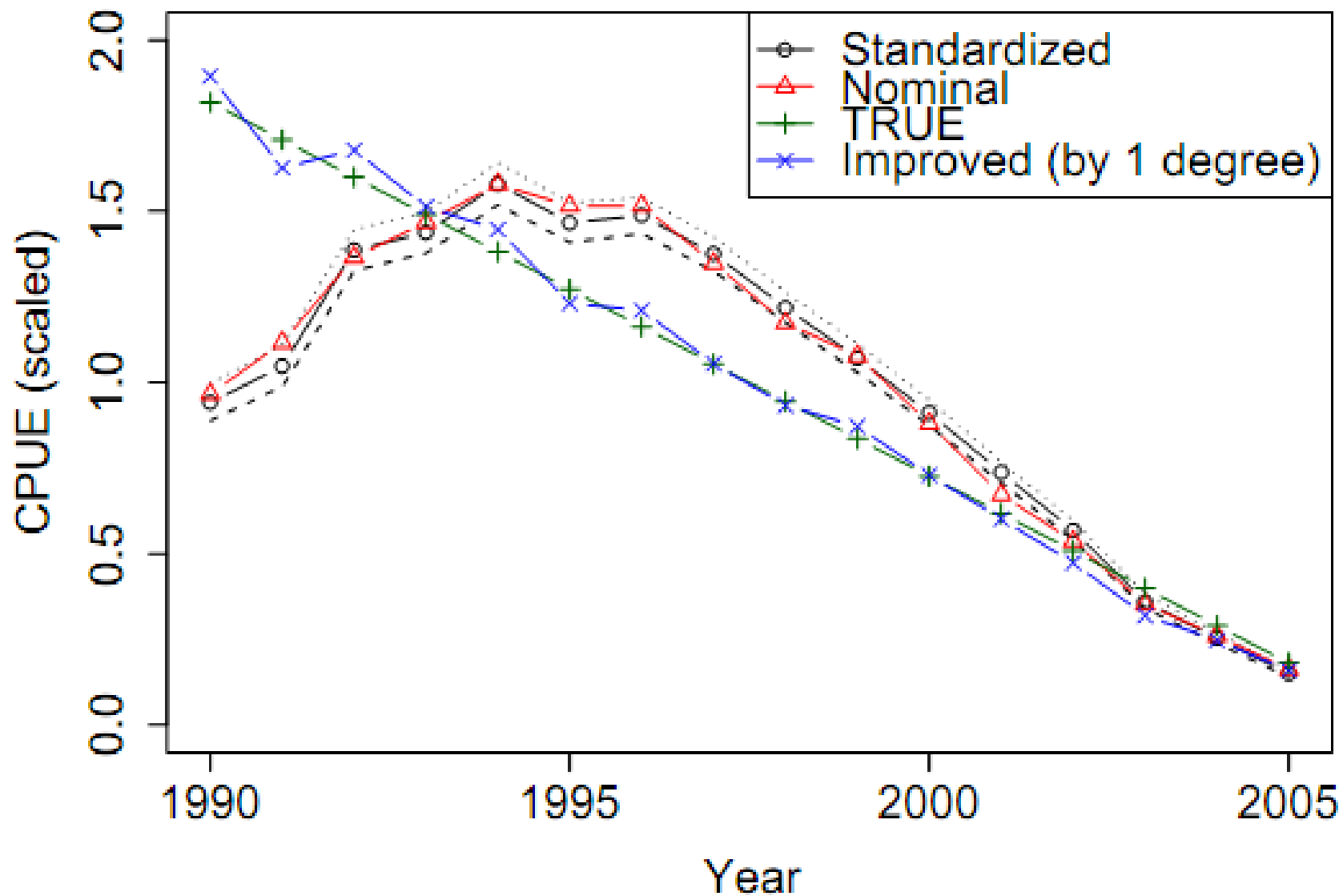
# 緯度・経度を効果にいった場合



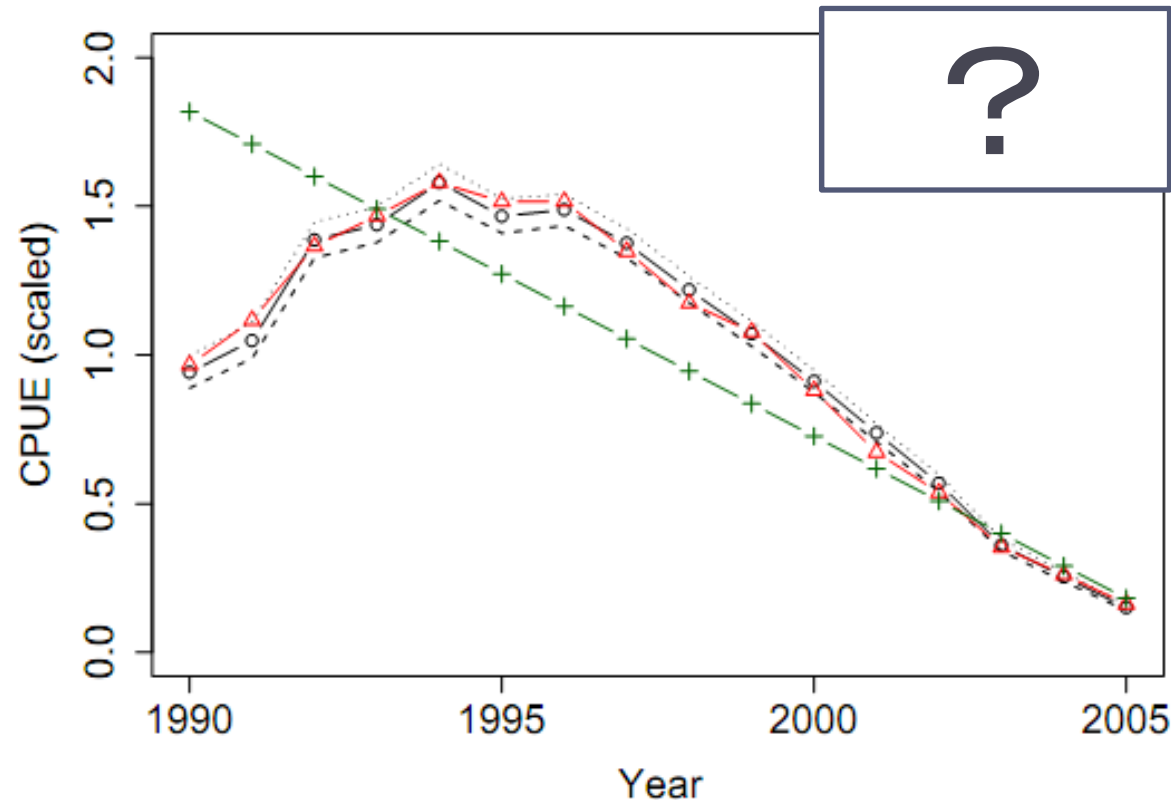
# 緯度・経度を効果にいった場合



# 緯度・経度を効果にいった場合



## 標準化CPUE: 解釈の難しさ



- 正解を知らずに、2つのトレンドのCPUEが出てきた時、正しいほうを選択できるだろうか？

# CPUE標準化の必要性

- はえ縄以外の様々な漁業データにおいても標準化CPUEは推定され、利用されている
- 国際資源評価で、標準化されていないCPUEが資源量指数として受け入れられることは基本的にない
  - CPUEの相対的な重要性が増加（統合モデル）

但し、標準化してさえいればいいというものではない。

- 標準化CPUEが本当に真の資源量を反映しているのか
- 操業分布や漁獲対象種・漁具の歴史的な変化による潜在的なバイアス
- GLMだけでない様々な統計モデル

# ご静聴ありがとうございました

## 参考文献

**LM一般の話:**「一般線形モデルによる生物科学のための現代統計学」Alan Grafen, Rosie Halis (著), 野間口謙太郎・野間口眞太郎(訳)

**CPUE標準化:** 庄野宏. 2008. 統計モデルとデータマイニング手法の水産資源解析への応用. 水研センター研報 **22**: 1-85.

**CPUE標準化:** Maunder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. *Fish. Res.* **70**: 141-159.

**宣伝:** Ichinokawa, M., and Brodziak, J. 2010. Using adaptive area stratification to standardize catch rates with application to North Pacific swordfish (*Xiphias gladius*). *Fish Res* **106(3)**: 249-260.