

# 生物資源解析学演習

## Lecture 5 回帰分析

東京海洋大学      北門   利英

## 回帰分析群

- 線形回帰分析(linear model) “lm”
- 分散分析(analysis of variance) “aov” , “anova”
- 非線形回帰分析(nonlinear model) “nls”
- 一般化線形モデル(generalized linear model) “glm”
- 一般化加法モデル(generalized additive model)  
“gam” in library(mgcv)

その他

- 樹形(樹木)モデル(tree-based model) “rpart”
- ニューラルネットワーク などなど

## 回帰分析群

	Distributional assumption	Regression component	R function
Normal linear model	Normal	Linear	"lm"
Normal nonlinear model	Normal	Nonlinear	"nls"
Generalized linear model (GLM)	Exponential family (Normal, Gamma, Binomial, Poisson etc)	Linear through "a link function"	"glm"
Nonparametric regression model	Normal	Nonparametric	"gam"
Generalized additive model (GAM)	Exponential family (Normal, Gamma, Binomial, Poisson etc)	Nonparametric through "a link function"	"gam"

## 非線形最小2乗法

求めたい関係式  $y = f(x; \theta)$

観測データ  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

データに対するモデル (加法誤差モデル)

$y_i = f(x_i; \theta) + \varepsilon_i$ ,  $\varepsilon_i$  は平均 0, 分散  $\sigma^2$  に従う誤差項

パラメータ  $\theta$  の求め方ー最小2乗基準ー

$$S(\theta) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - f(x_i; \theta))^2 \rightarrow \min$$

# 成長式の推定

## #データの読み込みと図示

```
Data <- read.csv("alfonsino.csv", header=T)
Age <- Data$Age
Length <- Data$Length
Sex <- as.numeric(Data$Sex)
mark <- c(19,5)
plot(Length~Age,pch=mark[Sex],xlim=c(0,25),ylim=c(0,50))
legend(18,20,pch=mark,legend=c("Female","Male"),bty="n")
```

## # von Bertalanffy 曲線の定義

```
growth.VB <- function(t,Linf,K,t0) Linf*(1.0-exp(-K*(t-t0)))
```

# 成長式の推定

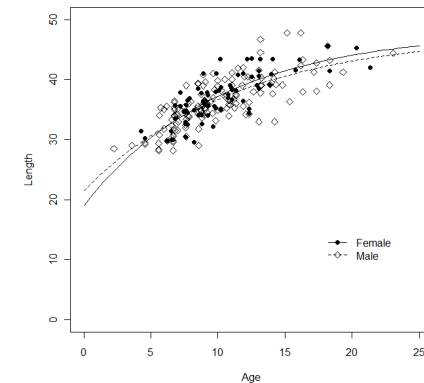
## # パラメータの推定(雌雄差あり)

```
start <- list(Linf=rep(max(Length),2),K=c(0.1,0.1),t0=c(0,0))
res.111 <- nls(Length~growth.VB(t=Age,Linf[Sex], K[Sex], t0[Sex]), start=start,
  trace=T)
summary(res.111)
confint(res.111)
```

```
taxis <- seq(0,25,0.5)
tlen <- length(taxis)
```

```
pred.female.111 <- predict(res.111, list(Age=taxis, Sex=rep(1,tlen)))
pred.male.111 <- predict(res.111, list(Age=taxis, Sex=rep(2,tlen)) )
```

```
plot(Length~Age,pch=mark[Sex],xlim=c(0,25),ylim=c(0,50))
points(taxis, pred.female.111, type="l", lty=1)
points(taxis, pred.male.111, type="l", lty=2)
legend(18,15,pch=mark,lty=c(1,2),legend=c("Female","Male"),bty="n")
```

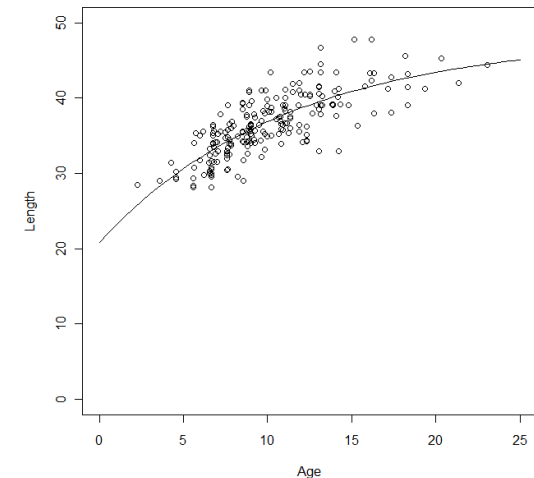


# 成長式の推定

## # パラメータの推定(雌雄差なし)

```
start <- list(Linf=max(Length),K=0.1,t0=0)
res.000 <- nls(Length~growth.VB(t=Age,Linf, K, t0), start=start, trace=T)
summary(res.000)
confint(res.000)

win.graph()
pred.000 <- predict(res.000, list(Age=taxis))
plot(Length~Age,xlim=c(0,25),ylim=c(0,50))
points(taxis, pred.000, type="l")
```



# 成長式の推定

## # モデル選択

AIC(res.111)

AIC(res.000)

## # 尤度比検定

# H0: 成長式に雌雄差なし

# H1: 成長式に雌雄差あり

```
lambda <- -2*(logLik(res.000)-logLik(res.111))
```

```
DF <- 7-4
```

```
if(lambda <= qchisq(0.95, DF))
```

```
  print("H0: accepted") else
```

```
  print("H0: rejected")
```



# Generalized Linear Models

## 正規線形モデル

- 確率分布：正規
- 回帰構造：線形

## 一般化線形モデル

- 確率分布：2項，ポアソン，ガンマなど  
正規分布以外の確率分布に対応
- 回帰構造：対数変換などの変換によっ  
て線形になればよい

# Generalized Linear Models

- 基本的には確率変数は独立と考える

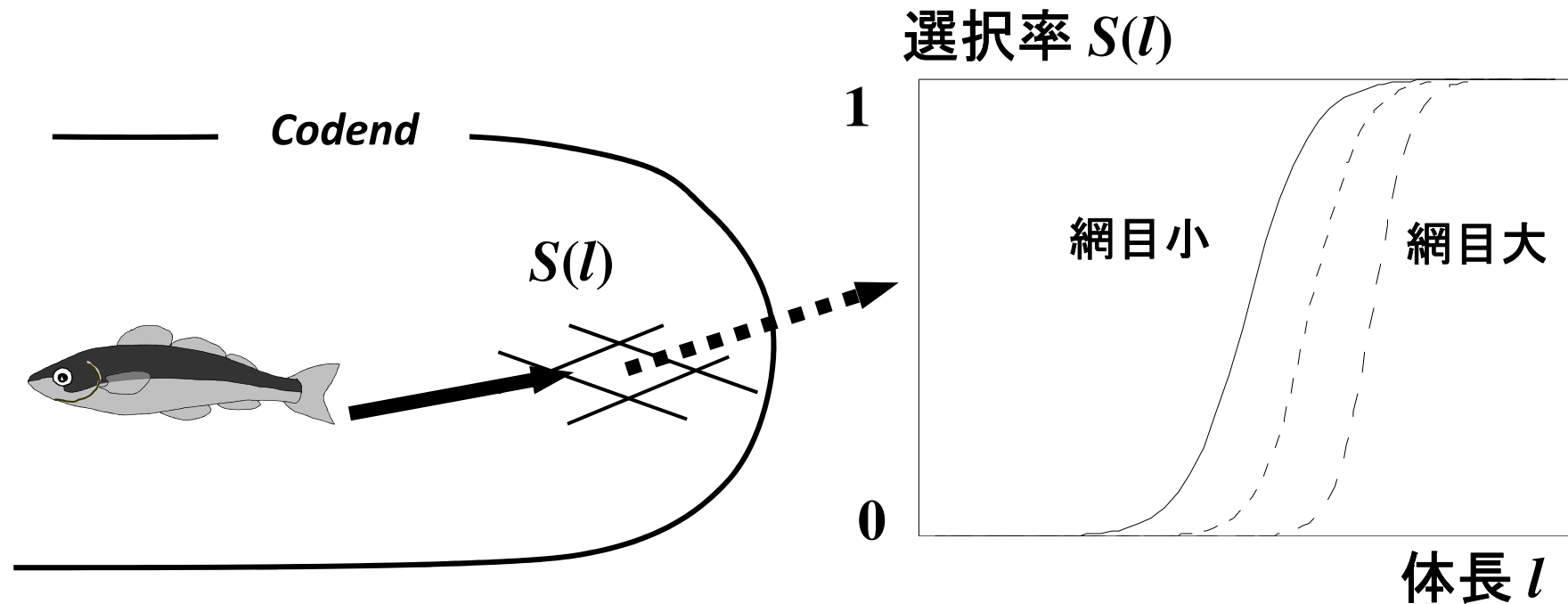
$$Y_i \sim f(y_i; \theta)$$

- 期待値パラメータを単調な関数で変換したときに，説明変数の1次式となる

$$\mu_i = E[Y_i]$$

$$g(\mu_i) = a + \sum_j b_j x_{ij} \text{ (} g \text{ を連結関数と呼ぶ)}$$

## Example : estimation of gear selectivity curves

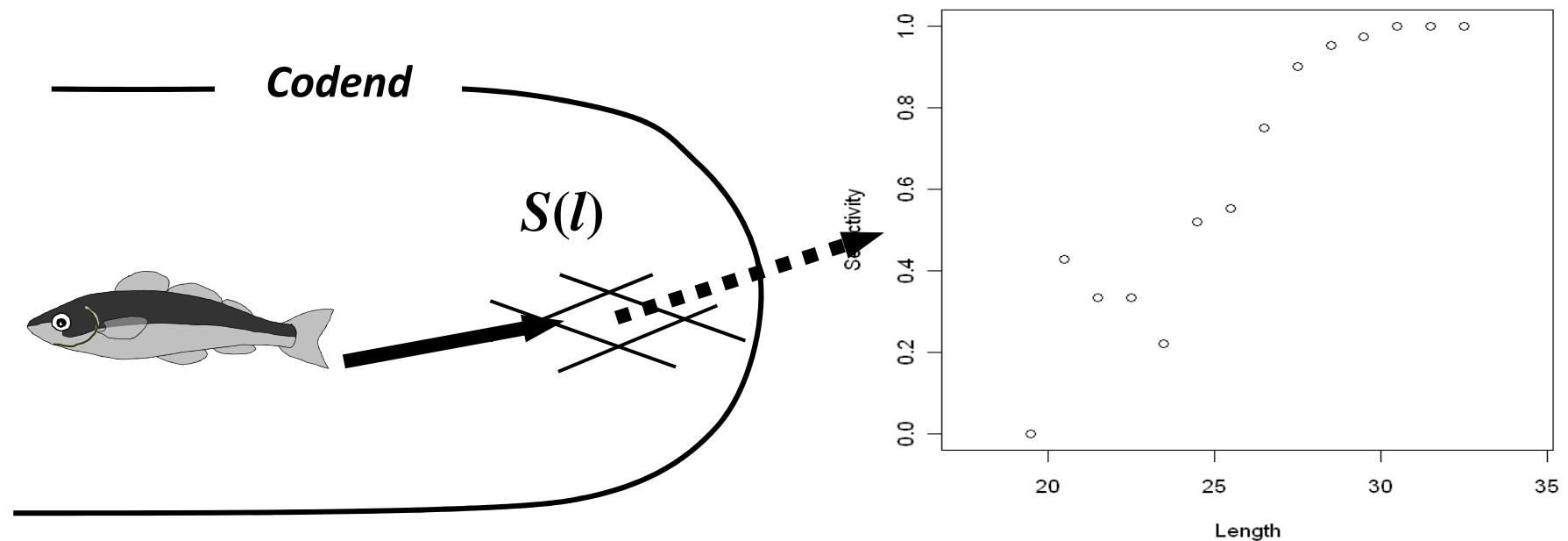


選択率＝網目に遭遇した魚が網に保持される確率

ロジスティック曲線

$$s(l) = \frac{\exp(a + bl)}{1 + \exp(a + bl)} \quad (a < 0, b > 0).$$

## Example : estimation of gear selectivity curves



選択率＝網目に遭遇した魚が網に保持される確率

ロジスティック曲線

$$s(l) = \frac{\exp(a + bl)}{1 + \exp(a + bl)} \quad (a < 0, b > 0).$$

# Rev: Estimation of a binomial parameter

## Binomial distribution

$$X_i \sim \text{Bin}(N_i, p) \quad (i = 1, \dots, n)$$

$$\Pr(X_i = x_i) = \binom{N_i}{x_i} p^{x_i} (1-p)^{N_i-x_i}$$

## The likelihood function

$$L(p) = \prod_{i=1}^n \binom{N_i}{x_i} p^{x_i} (1-p)^{N_i-x_i}$$

## The log-likelihood function

$$l(p) = \log L(p) = \sum_{i=1}^n \log \binom{N_i}{x_i} + \sum_{i=1}^n [x_i \log p + (N_i - x_i) \log(1-p)]$$

# Rev: Estimation of binomial parameters

## Binomial distribution

$$X_i \sim \text{Bin}(N_i, p_i) \quad (i = 1, \dots, n)$$

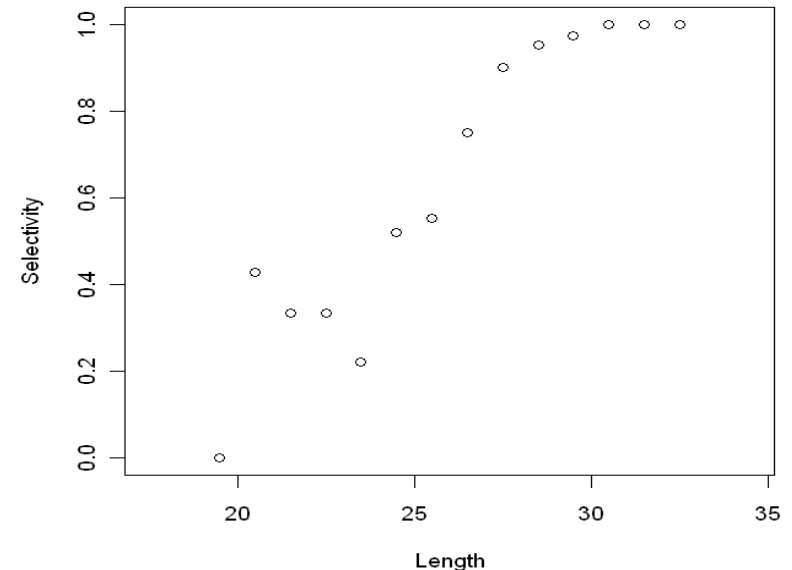
$$\Pr(X_i = x_i) = \binom{N_i}{x_i} p_i^{x_i} (1 - p_i)^{N_i - x_i}$$

## The likelihood function

$$L(p_1, \dots, p_n) = \prod_{i=1}^n \binom{N_i}{x_i} p_i^{x_i} (1 - p_i)^{N_i - x_i}$$

## The log-likelihood function

$$l(p_1, \dots, p_n) = \log L(p_1, \dots, p_n) = \sum_{i=1}^n \log \binom{N_i}{x_i} + \sum_{i=1}^n [x_i \log p_i + (N_i - x_i) \log(1 - p_i)]$$



# Estimation of **regression coefficients** in a logistic curve

## Binomial distribution

$$X_i \sim \text{Bin}(N_i, s(l_i)) \quad (i = 1, \dots, n)$$

$$\Pr(X_i = x_i) = \binom{N_i}{x_i} s(l_i)^{x_i} (1 - s(l_i))^{N_i - x_i}$$

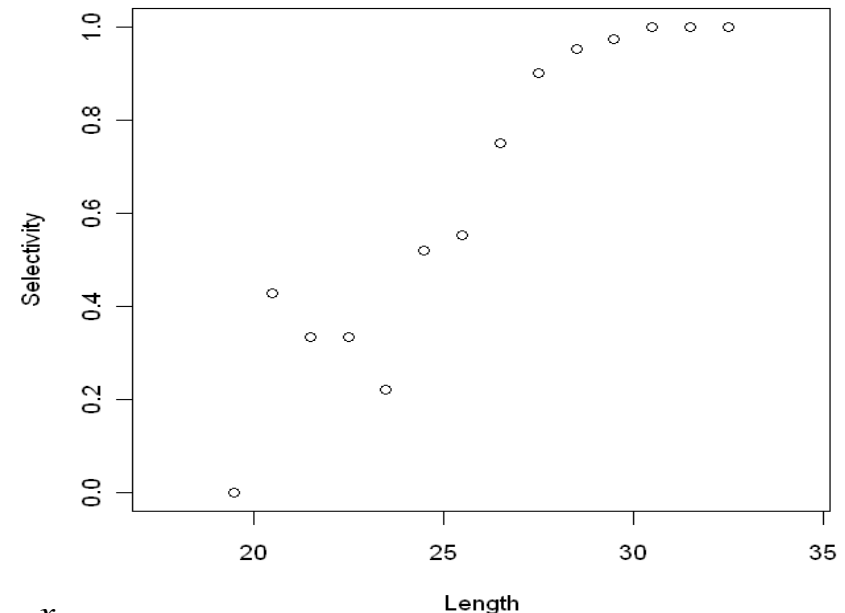
$$s(l) = \frac{e^{a+bl}}{1 + e^{a+bl}}$$

## The likelihood function

$$L(a, b) = \prod_{i=1}^n \binom{N_i}{x_i} s(l_i)^{x_i} (1 - s(l_i))^{N_i - x_i}$$

## The log-likelihood function

$$l(a, b) = \log L(a, b) = \sum_{i=1}^n \log \binom{N_i}{x_i} + \sum_{i=1}^n [x_i \log s(l_i) + (N_i - x_i) \log(1 - s(l_i))]$$



## Example : estimation of gear selectivity curves

```
#Boerema(1956)
```

```
Length<-seq(19.5, 32.5)
```

```
Codend<-c(0,3,3,5,4,13,26,27,46,62,38,29,28,19)
```

```
Cover<-c(5,4,6,10,14,12,21,9,5,3,1,0,0,0)
```

```
n<- Codend + Cover
```

```
res.logit<-glm(cbind(Codend, Cover)~Length, family=binomial)
```

```
summary(res.logit)
```



## Example : estimation of gear selectivity curves

#GLMでは信頼区間が比較的簡単に求まる(関数形が限られているので)

pred.se<-predict (res.logit, se.fit=TRUE) #線形項の予測

#線形項(a+bl)の信頼区間

Lower <- pred.se\$fit - qnorm(0.975, 0, 1)\* pred.se\$se.fit

Upper <- pred.se\$fit + qnorm(0.975, 0, 1)\* pred.se\$se.fit

#ロジスティックの信頼区間に変換

logit.t<-function(x){ 1/(1+exp(-x)) }

pred.logit<- sapply(pred.se\$fit, logit.t)

Lower.logit <- sapply(Lower, logit.t)

Upper.logit<- sapply(Upper, logit.t)

pcon.logit<-cbind(pred.logit, Lower.logit, Upper.logit)

plot(Length, Codend/n, xlim=c(18,35), ylim=c(0,1))

matlines(Length, pcon.logit, lty=c(1,2,2), col="blue", lwd=1.5)

## Example : estimation of gear selectivity curves

### 補足

#注1: 先ほどの例は少しサボりましたが,

```
newx<-seq(18,33,0.5)
```

```
pred.se<-predict (res.sel, list(Length=newx),se.fit=TRUE)
```

```
...
```

```
matlines(newx, pcon, lty=c(1,2,2), col=c("blue","red","red"))
```

とした方が綺麗に図が描けます

#注2: もしロジスティック関数の予測値を直接求めたいだけなら

```
pred.se<-predict (res.sel, type="response", se.fit=TRUE)
```

を利用

## Example : estimation of gear selectivity curves

①logistic関数以外の利用(例えば probit モデル)

```
res.probit<-glm(cbind(Codend, Cover)~Length, family=binomial(link=probit))
```

```
pred.se<-predict (res.probit, se.fit=TRUE) #線形項の予測
```

```
#線形項(a+bl)の信頼区間
```

```
Lower <- pred.se$fit - qnorm(0.975, 0, 1)* pred.se$se.fit
```

```
Upper <- pred.se$fit + qnorm(0.975, 0, 1)* pred.se$se.fit
```

```
#ロジスティックの信頼区間に変換
```

```
probit.t<-function(x){ pnorm(x,0,1)}
```

```
pred.probit<- sapply(pred.se$fit, probit.t)
```

```
Lower.probit <- sapply(Lower, probit.t)
```

```
Upper.probit<- sapply(Upper, probit.t)
```

```
pcon.probit<-cbind(pred.probit, Lower.probit, Upper.probit)
```

```
matlines(Length, pcon.probit, lty=c(1,2,2), col="red", lwd=2)
```

```
legend(28, 0.3, c("logit","probit "), lty=c(1,1), col=c("blue", "red"), lwd=1.5,  
      bty="n")
```

```
AIC(res.logit)
```

```
AIC(res.probit)
```

## Example : estimation of gear selectivity curves

②ポアソン分布に従う変数に log-linear モデルを仮定する場合は

```
glm(y~x, family="poisson")
```

で解析可能.

③glmmを利用すればランダム効果モデルを含む解析も可能  
(省略します)

# GAM(1)

ちょっと資源解析的な背景を忘れて単にモデルとして

$$Y_i \sim N(s(x_i), \sigma^2) \quad (i = 1, \dots, n)$$

$s(x)$ :ノンパラメトリックな関数を仮定

## #模擬データ

```
x<-seq(1,100)
error<-rnorm(100, 0, 1)
y<-log(x) + 1/x + sin(x*pi/20) + error
plot(x,y)
library(mgcv)
res.gam1 <- gam(y~s(x), family="gaussian"); plot(res.gam1)
pred.se<-predict(res.gam1, type="response", se.fit=TRUE)
conp<-
  cbind(pred.se$fit, pred.se$fit-1.96*pred.se$se.fit, pred.se$fit+1.96*pred.se$se.fit )
plot(x,y); matlines(x, conp, lty=c(1,2,2))
```

## GAM(2)

続いて資源解析的な背景を忘れて単にモデルとして

$$X_i \sim \text{Bin}(N_i, s(l_i)) \quad (i = 1, \dots, n)$$

$$\Pr(X_i = x_i) = \binom{N_i}{x_i} s(l_i)^{x_i} (1 - s(l_i))^{N_i - x_i}$$

$s(l)$ :ノンパラメトリックな関数を仮定

```
library(mgcv)
```

```
res.gam <- gam(cbind(Codend,Cover)~s(Length), family="binomial")
```

以下glmと同様に処理可能

## 5. 番外編

## その他の配布物

- Rによる最適化と最尤推定
- プロダクションモデル(最尤推定)
- プロダクションモデル(ベイズ推定)
- Rを用いたシミュレーション