Приложение 9

## MASE за 5 лет по ортогональной регрессии в западной части Камчатки

Загрузим данные в pinks и удалим последние 5 лет в pinksTrain, сохранив их для тестирования в pinksTest.

pinks <- readxl::read\_excel('data.xlsx', sheet = 'Запад')
pinks <- transform(pinks, S=exp(LnS), R=exp(LnR))
pinksTrain = pinks[1:27,]
pinksTest = pinks[28:32,]

Далее всё как и прежде.

library(FSA)

## ## FSA v0.9.5. See citation('FSA') if used in publication.
## ## Run fishR() for related website and fishR('IFAR') for related book.

rckr <- srFuns('Ricker')

Получим стартовые значения для оптимизатора.

(svR <- srStarts(R ~ S, data=pinksTrain, type='Ricker'))

## $a
## [1] 3.81754
##
## $b
## [1] 0.01615104

Загрузим пакет для ортогональной нелинейной регрессии

library(onls)

## Loading required package: minpack.lm

ЗДЕСЬ УЖЕ БУДЕТ ОРТОГОНАЛЬНАЯ регрессия из pinksTrain!

srR <- onls(LnR~log(rckr(S,a,b)),data=pinksTrain,start=svR)

## Obtaining starting parameters from ordinary NLS...
## Passed...
## Conditions for `info = 1' and `info = 2' both hold.
## Optimizing orthogonal NLS...
## Passed... Conditions for `info = 1' and `info = 2' both hold.

print(srR)

## Nonlinear orthogonal regression model
## model: LnR ~ log(rckr(S, a, b))
## data: pinksTrain
## a b
## 4.96245 0.02182
## vertical residual sum-of-squares: 187.6
## orthogonal residual sum-of-squares: 21.91
## PASSED: 27 out of 27 fitted points are orthogonal.
##
## Number of iterations to convergence: 4
## Achieved convergence tolerance: 1.49e-08

График

plot(srR)



Статистика

summary(srR)

##
## Formula: LnR ~ log(rckr(S, a, b))
##
## Parameters:
## Estimate Std. Error t value Pr(>|t|)
## a 4.96245 3.51458 1.412 0.170
## b 0.02182 0.02456 0.888 0.383
##
## Residual standard error of vertical distances: 2.74 on 25 degrees of freedom
## Residual standard error of orthogonal distances: 0.9361 on 25 degrees of freedom
##
## Number of iterations to convergence: 4
## Achieved convergence tolerance: 1.49e-08

confint(srR)

## 2.5% 97.5%
## a 2.479729778 10.16671355
## b 0.001218078 0.04266158

Доверительный интервал b НЕ пересекает 0, но его ошибка выше ожидания. Нужна проверка значимости фактора плотности (т.к. он (b) может быть лишним).

ind <- srFuns('independence')
svI <- srStarts(R~S,data=pinksTrain,type='independence')
srI <- onls(LnR~log(ind(S,a)),data=pinksTrain,start=svI)

## Obtaining starting parameters from ordinary NLS...
## Passed...
## Relative error in the sum of squares is at most `ftol'.
## Optimizing orthogonal NLS...
## Passed... Relative error in the sum of squares is at most `ftol'.

# Отношение правдоподобия
lrt(srI,com=srR)

## Model 1: LnR ~ log(ind(S, a))
## Model A: LnR ~ log(rckr(S, a, b))
##
## DfO logLikO DfA logLikA Df logLik Chisq Pr(>Chisq)
## 1vA 26 -67.2139 25 -64.4824 1 -2.7315 5.4629 0.01942 \*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Выходит, что фактор плотности имеет высокую статистическую значимость (p < 0.02).

Визуализируем связь по Рикеру.

x <- seq(min(pinksTrain$S)\*.5, max(pinksTrain$S)\*1.5,length.out=199) # many S for prediction
pR <- rckr(x,a=coef(srR)) # прогнозируемый возврат по Рикеру, R
pI <- ind(x,a=coef(srI)) # прогнозируемый возврат без фактора плотности (b), R
plot(R ~ S,data=pinksTrain,col='white',ylab='Возврат', xlab='Нерест')
points(R~S,data=pinksTrain,pch=19,col=rgb(0,0,0,1/2))
lines(pR~x,lwd=2, col='green')
lines(pI~x,lwd=2, col='red')
legend('topleft', legend=c('Рикер', 'Без фактора плотности'),
 col = c('green','red'), lty=1)



Модель без фактора плотности с коэффициентами от ортогональной регрессии.

summary(srI)

##
## Formula: LnR ~ log(ind(S, a))
##
## Parameters:
## Estimate Std. Error t value Pr(>|t|)
## a 2.786 1.594 1.748 0.0922 .
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error of vertical distances: 2.972 on 26 degrees of freedom
## Residual standard error of orthogonal distances: 1.001 on 26 degrees of freedom
##
## Number of iterations to convergence: 2
## Achieved convergence tolerance: 1.49e-08

confint(srI)

## 2.5% 97.5%
## 1.762755 4.429939

plot(srI)



# MASE - Mean absolute scaled error

Оценим шкалированную ошибку на последних 5 годах в модели без фактора плотности

library(yardstick)
pinksTrain$RhatInd <- ind(pinksTrain$S,a=coef(srI))
pinksTrain$LaggedR = c(NA,pinksTrain$R[1:26])
MAEnaive = mae(pinksTrain, truth=R, estimate=LaggedR)$`.estimate`
MAEtrainInd = mae(pinksTrain, truth=R, estimate=RhatInd)$`.estimate`
pinksTest$RhatInd <- ind(pinksTest$S,a=coef(srI))
MASE\_Ind = mase(pinksTest, truth=R, estimate=RhatInd, mae\_train = MAEnaive)$`.estimate`
paste('MASE Без фактора плотности в тесте =', round(MASE\_Ind,3))

## [1] "MASE Без фактора плотности в тесте = 1.157"

Оценим шкалированную ошибку на последних 5 годах в модели Рикера

pinksTrain$RhatRic <- rckr(pinksTrain$S,a=coef(srR))
MAEtrainRic = mae(pinksTrain, truth=R, estimate=RhatRic)$`.estimate`
pinksTest$RhatRic <- rckr(pinksTest$S,a=coef(srR))
MASE\_Ric = mase(pinksTest, truth=R, estimate=RhatRic, mae\_train = MAEnaive)$`.estimate`
paste('MASE в модели Рикера в тесте =', round(MASE\_Ric,3))

## [1] "MASE в модели Рикера в тесте = 1.031"

Итак, по MASE классическая модель Рикера на западе Камчатки прогнозирует не лучше наивного прогноза (MASE>1) и без фаткора плотности хуже (MASE > 1).

Посмотрим ещё на ошибки в обучении (здесь на pinksTrain). Они могут быть не так сильно отличными, например, MAE - mean absolute error.

paste('MAE Без фактора плотности в обучении на pinksTrain =', round(MAEtrainInd,3))

## [1] "MAE Без фактора плотности в обучении на pinksTrain = 36.398"

paste('MAE в модели Рикера в обучении на pinksTrain =', round(MAEtrainRic,3))

## [1] "MAE в модели Рикера в обучении на pinksTrain = 30.789"

Здесь МАЕ у модели Рикера ниже, чем в модели без фактора плотности.