Приложение 5

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

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

options(OutDec= ',')
pinks <- readxl::read\_excel('data.xlsx', sheet = 'Восток')
pinks <- transform(pinks, LnR=log(R))
pinksTrain = pinks[1:27,]
pinksTest = pinks[28:32,]

Далее всё как и в Приложении 3.

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] 2,92918
##
## $b
## [1] 0,006774296

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
## 2,849746 0,006147
## vertical residual sum-of-squares: 53,53
## orthogonal residual sum-of-squares: 18,51
## PASSED: 27 out of 27 fitted points are orthogonal.
##
## Number of iterations to convergence: 3
## 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 2,849746 1,197624 2,379 0,0253 \*
## b 0,006147 0,011999 0,512 0,6130
## ---
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1
##
## Residual standard error of vertical distances: 1,463 on 25 degrees of freedom
## Residual standard error of orthogonal distances: 0,8604 on 25 degrees of freedom
##
## Number of iterations to convergence: 3
## Achieved convergence tolerance: 1,49e-08

confint(srR)

## 2,5% 97,5%
## a 1,689783744 4,81187669
## b -0,008586894 0,02087425

Ошибка b выше его ожидания, поэтому его доверительный интервал пересекает 0. 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 -48,7780 25 -47,5508 1 -1,2272 2,4544 0,1172

Выходит, что фактор плотности имеет низкую статистическую значимость (p > 0.1).

Визуализируем связь по Рикеру.

x <- seq(min(pinksTrain$S)\*.5, max(pinksTrain$S)\*1.5,length.out=199) #
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)



Очевидно, что значимых преимуществ у модели по Рикеру нет.

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 2,4181 0,6988 3,46 0,00188 \*\*
## ---
## Signif. codes: 0 '\*\*\*' 0,001 '\*\*' 0,01 '\*' 0,05 '.' 0,1 ' ' 1
##
## Residual standard error of vertical distances: 1,502 on 26 degrees of freedom
## Residual standard error of orthogonal distances: 0,8559 on 26 degrees of freedom
##
## Number of iterations to convergence: 3
## Achieved convergence tolerance: 1,49e-08

confint(srI)

## 2,5% 97,5%
## 1,716760 3,408383

plot(srI)



# MASE - Mean absolute scaled error

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

library(yardstick)
# mean absolute error
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 = 0,599"

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 = 1,167"

Итак, по MASE классическая модель Рикера на востоке Камчатки прогнозирует хуже наивного прогноза (MASE>1), а без фаткора плотности лучше (MASE < 1).

Поэтому важнее смотреть MASE OOB, т.к. ошибки в обучении (здесь на pinksTrain) могут быть не так сильно отличными, например, MAE - mean absolute error и привести к неверным выводам.

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

## [1] "MAE Без фактора в обучении на pinksTrain = 28,225"

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

## [1] "MAE в модели Рикера в обучении на pinksTrain = 28,715"

Хоть МАЕ и чуть выше у модели Рикера была, но не так наглядно, как MASE в тесте.