Title

Firstname Lastname
Affiliation

Abstract

—!!!—an abstract is required—!!!—

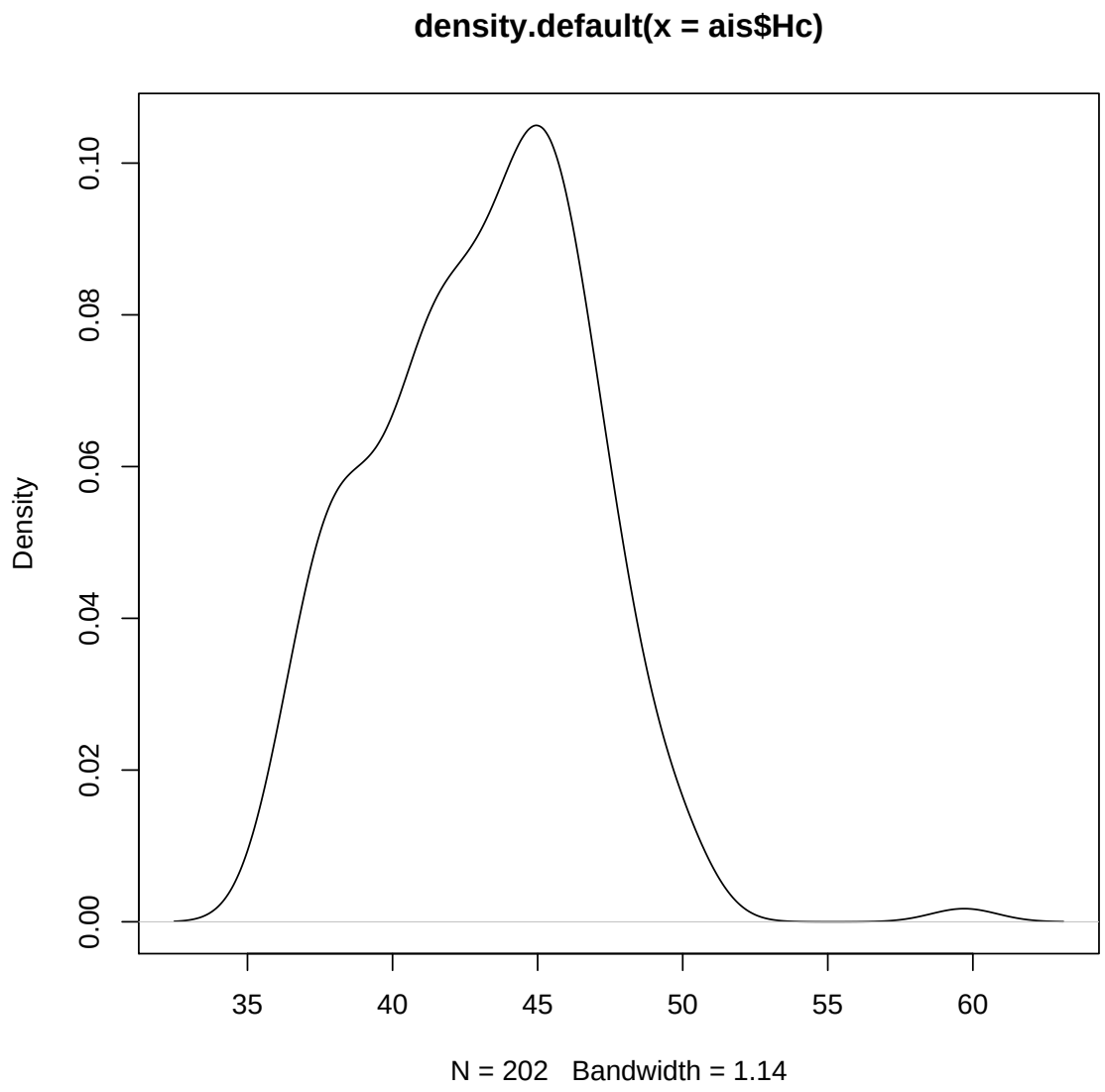
Keywords: —!!!—at least one keyword is required—!!!—.

Illustrates focused model comparison for a novel class of models.

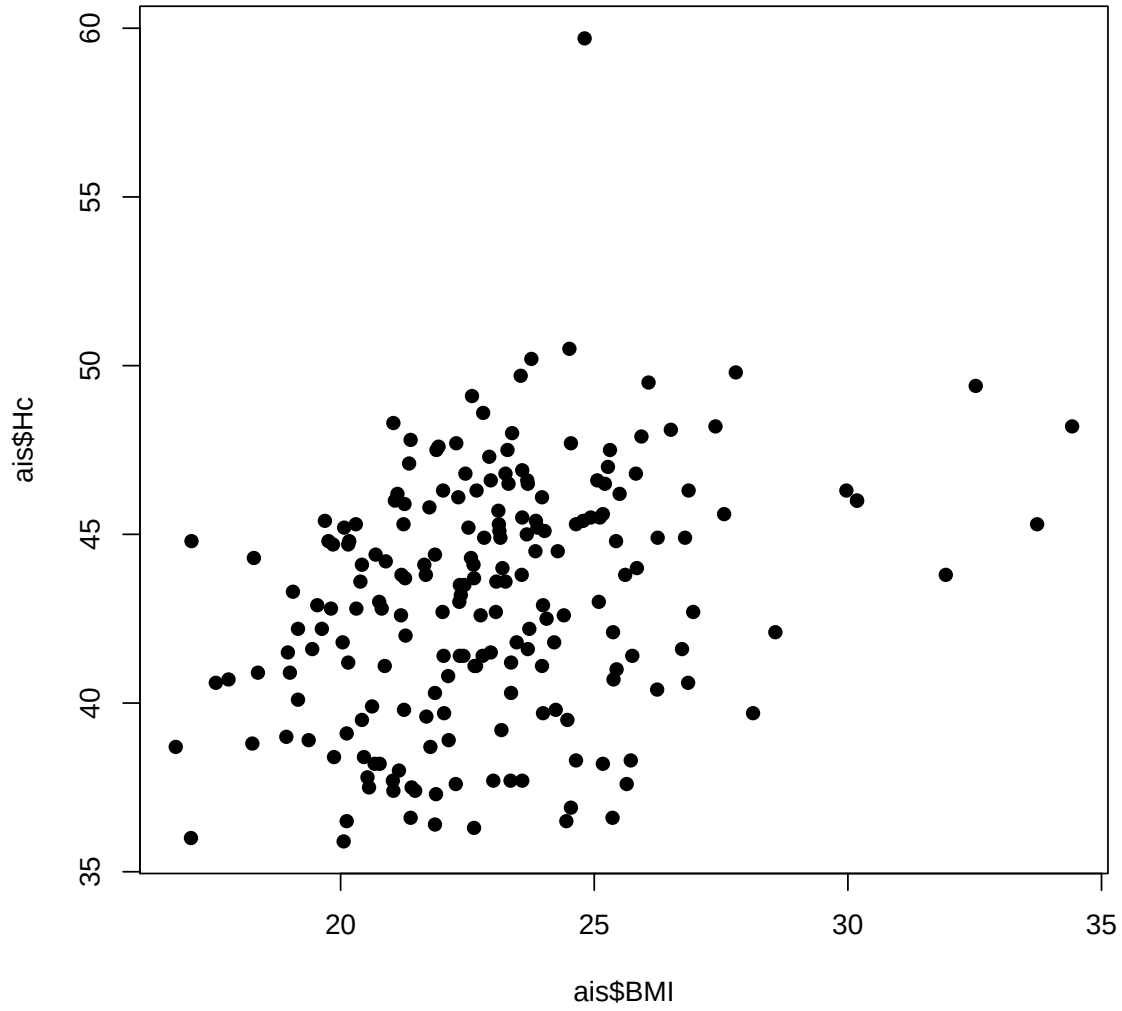
This needs narrow model pars to be in the middle!

Mixture of covariate selection and flexibility selection. Covariate coefficients are subset of model parameters

```
if (!require("sn")) stop("The `sn` package should be installed to run code in this vignette")
data(ais)
plot(density(ais$Hc))
```




```
plot(ais$BMI, ais$Hc, pch=19)
```




```
ldsnorm <- function(x, mean, sd, skew){
    log(skew) + (skew-1)*pnorm(x, mean, sd, log=TRUE) + dnorm(x, mean, sd, log=TRUE)
}

mloglik <- function(b0, b1, sd, skew){
    ret <- -sum(ldsnorm(ais$Hc, b0 + b1*ais$BMI, sd, skew))
    ret
}

snlm <- function(reg=FALSE, skew=FALSE){
    inilm <- lm(Hc ~ BMI, data=ais)
    ini <- c(coef(inilm), log(summary(inilm)$sigma), 0)
    if (!reg && !skew){
```

```
        ini <- ini[c(1,3)]
        fn <- function(par) mloglik(par[1], 0, exp(par[2]), 1)
    } else if (reg && !skew){
        ini <- ini[c(1,2,3)]
        fn <- function(par) mloglik(par[1], par[2], exp(par[3]), 1)
    } else if (!reg && skew){
        ini <- ini[c(1,3,4)]
        fn <- function(par) mloglik(par[1], 0, exp(par[2]), exp(par[3]))
    } else if (reg && skew){
        fn <- function(par) mloglik(par[1], par[2], exp(par[3]), exp(par[4]))
    }
    opt <- nlm(fn, ini, hessian=TRUE)
    vcov <- solve(opt$hessian)
    list(loglik=-opt$minimum, est=opt$estimate, vcov=vcov, nobs=nrow(ais))
}
```

Four models

1. constant mean, constant variance, standard normal error
2. linear regression, no skewness ie standard normal error
3. no regression term, skewed error
4. linear regression term, skewed error

```
(mod1 <- snlm(reg=FALSE, skew=FALSE))

## Warning in nlm(fn, ini, hessian = TRUE): NA/Inf replaced by maximum positive
value

## $loglik
## [1] -548
##
## $est
## [1] 43.1  1.3
##
## $vcov
##          [,1]     [,2]
## [1,] 6.61e-02 1.06e-05
## [2,] 1.06e-05 2.48e-03
##
## $nobs
## [1] 202

(mod2 <- snlm(reg=TRUE, skew=FALSE))

## $loglik
## [1] -537
```

```
##
## $est
## [1] 33.68  0.41  1.24
##
## $vcov
##           [,1]      [,2]      [,3]
## [1,]  3.888463 -1.67e-01  1.74e-04
## [2,] -0.166805  7.27e-03 -7.22e-06
## [3,]  0.000174 -7.22e-06  2.48e-03
##
## $nobs
## [1] 202

(mod3 <- snlm(reg=FALSE, skew=TRUE))

## Warning in nlm(fn, ini, hessian = TRUE): NA/Inf replaced by maximum positive
value

## $loglik
## [1] -548
##
## $est
## [1] 39.659  1.547  0.929
##
## $vcov
##        [,1]    [,2]   [,3]
## [1,] 15.727 -0.9626 -3.823
## [2,] -0.963  0.0609  0.234
## [3,] -3.823  0.2335  0.934
##
## $nobs
## [1] 202

(mod4 <- snlm(reg=TRUE, skew=TRUE))

## $loglik
## [1] -537
##
## $est
## [1] 31.542  0.404  1.428  0.667
##
## $vcov
##        [,1]     [,2]     [,3]     [,4]
## [1,] 12.889 -0.14488 -0.68685 -2.58335
## [2,] -0.145  0.00715 -0.00138 -0.00521
## [3,] -0.687 -0.00138  0.05380  0.19422
```

```
## [4,] -2.583 -0.00521  0.19422  0.73537
##
## $nobs
## [1] 202
```

0.1. Mean at particular cov value, shouldn't need skewness to be known.

```
focus <- function(par, X){
    X %*% par[1:2]
}

fns <- list(coef=function(x)x$est,
            vcov=function(x)x$vcov,
            nobs=function(x)x$nobs)

med.bmi <- rbind(male=c(1, 23.56), female=c(1, 21.82))

inds <- rbind(c(1,0,1), c(1,1,1))
fic(mod2, inds=inds, inds0=c(1,0,1), fns=fns, focus=focus, X=med.bmi,
    sub=list(mod1, mod2))

##     vals mods     FIC  rmse rmse.adj   bias bias.adj    se focus
## 1   male    1   12.39 0.343    0.347 -0.248   -0.248 0.244  43.1
## 4   male    2    1.07 0.249    0.249  0.000    0.000 0.249  43.3
## 2 female    1   43.80 0.517    0.525  0.466    0.466 0.244  43.1
## 5 female    2    3.79 0.262    0.262  0.000    0.000 0.262  42.6
## 3    ave    1    2.40 0.266    9.217  0.109    9.214 0.244  43.1
## 6    ave    2 1550.02 2.781    1.974  0.000    0.000 1.974  43.0
```

models 3 and 4 have worse FIC, rmse as expected FIC values here near those in book but not exactly

0.2. Median at particular cov value, skewness may affect this. should pick biggest model.

0.3. Third central moment – pick skewed model if degree of skewness sufficiently big

todo should the focus function (mean) involve the error term for skew models?

```
inds <- rbind(c(1,0,1,0), c(1,1,1,0), c(1,0,1,1), c(1,1,1,1))
```