**Preguntes tipus Examen**

1. **Estimació No paramètrica de la densitat (Teoria)**
2. Explica detalladament quin és l’efecte de l’amplada de banda o de finestra sobre l’error quadràtic mig en l’estimació no paramètrica de la densitat efectuada amb un kernel gaussià.
   1. Definir l’error quadràtic mig en aquest contexte
   2. Definir com varia un kernel en funció de l’amplada de banda
   3. Determinar que passa amb l’error quadràtic mig per amplades de banda petites (menys obervacions) i que passa amb amplades de bandes grans (més observacions)
3. Explica detalladament com generar una variable aleatòria a partir d’una estimació no paramètrica de la densitat amb nucli gaussià i valors uniformes entre (0,1). Aplicat a bootstrap, quin nom rep aquest mètode?

(aquí estaria fer una estimació no paramètrica de la densitat d’una normal i comparar que és equivalent i veure com canvia respecte altres mètodes)

1. En l’estimació no paramètrica de la densitat efectuada a partir d’un histograma, explica detalladament quins elements s’han d’estimar/calcular i com aquests poden afectar a la seva forma.
2. **Estimació No paramètrica de la densitat (R)**
3. Indica dues diferències en la utilització de la comanda density o sm.density amb l’R per determinar-ne l’estimació no paramètrica de la densitat
   1. Estimació amb gran varietat de kernels = density
   2. Estimació amb kernel gaussià sm.density
   3. El suport on s’avaluen en density fa referència només a talls que no coincideixen amb la mostra, a sm.density conté ambdues formes
   4. Density es pot fer servir només pel cas univariant, sm.density permet treballar fins a 3 dimensions.
4. De les figures que es mostren a continuació; determineu amb quina opció us quedaríeu i per que (diferències en bandwidth i en kernel)

par(mfrow=c(2,2))

d1<-density(faithful$waiting, kernel="rectangular")

plot(d1, main="K.Rectangular")

d1<-density(faithful$waiting, kernel="rectangular", bw=2)

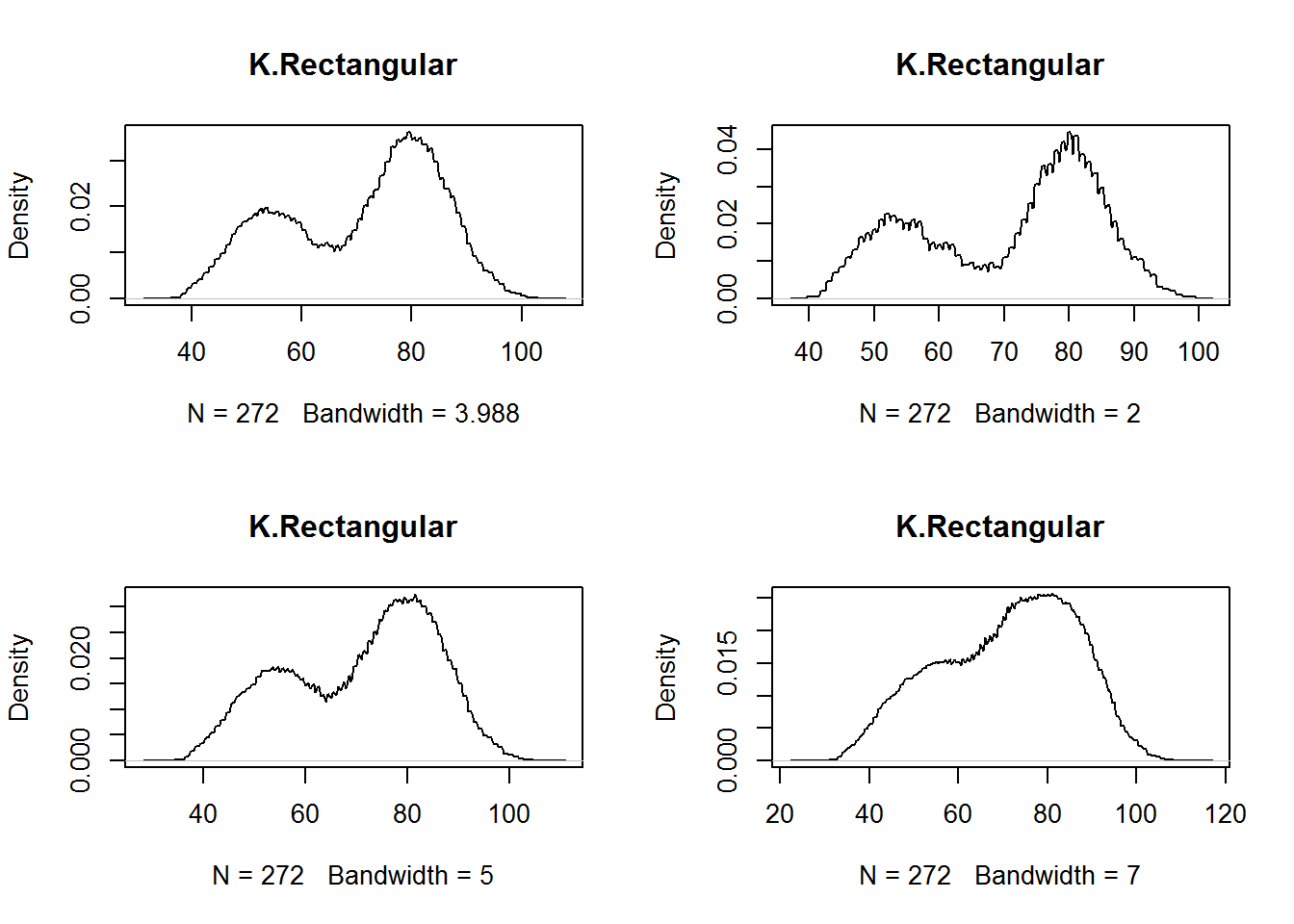
plot(d1, main="K.Rectangular")

d1<-density(faithful$waiting, kernel="rectangular", bw=5)

plot(d1, main="K.Rectangular")

d1<-density(faithful$waiting, kernel="rectangular", bw=7)

plot(d1, main="K.Rectangular")



par(mfrow=c(2,2))

d1<-density(faithful$waiting, kernel="epanechnikov")

plot(d1, main="K.Epanechnikov")

d1<-density(faithful$waiting, kernel="epanechnikov", bw=2)

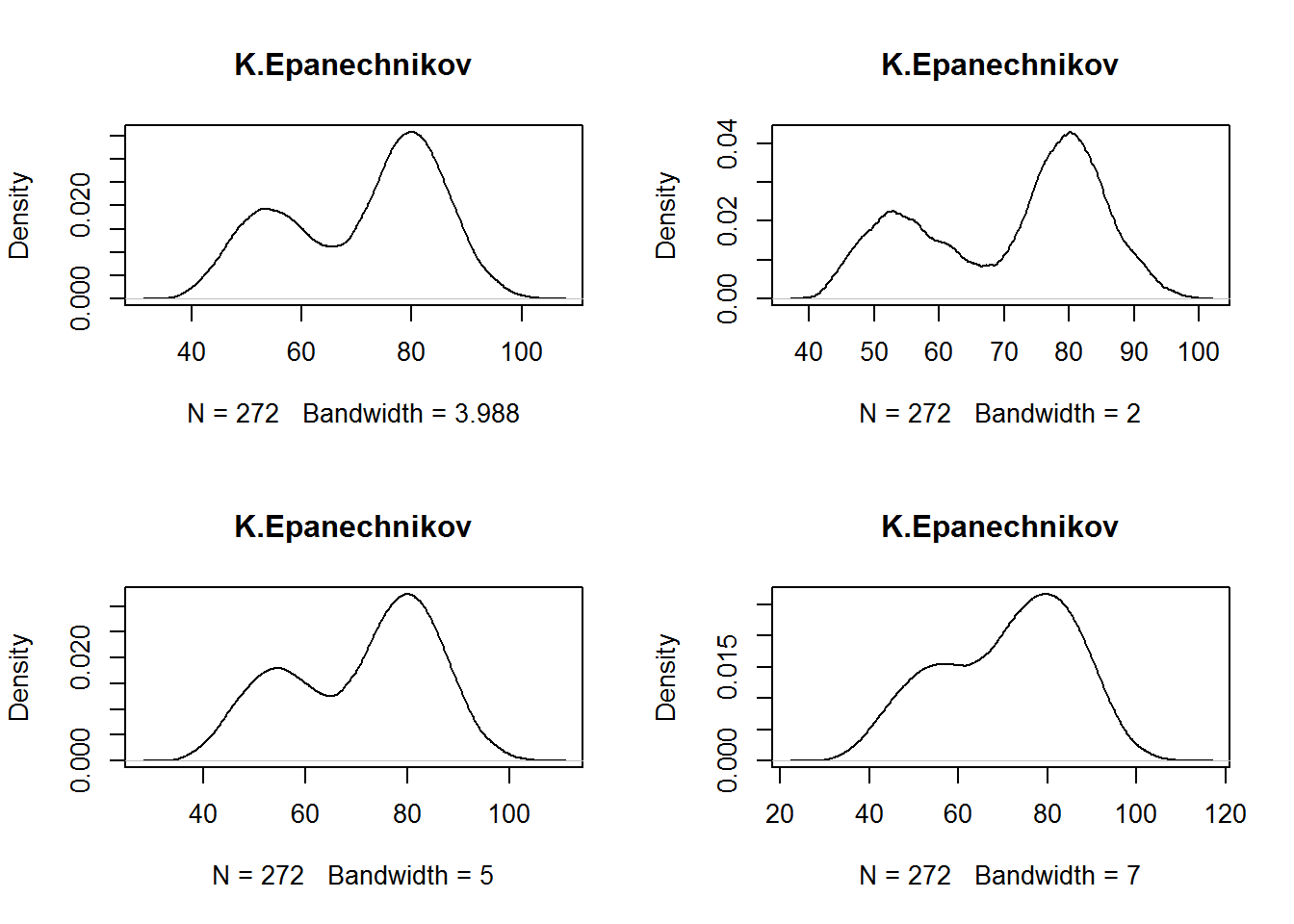
plot(d1, main="K.Epanechnikov")

d1<-density(faithful$waiting, kernel="epanechnikov", bw=5)

plot(d1, main="K.Epanechnikov")

d1<-density(faithful$waiting, kernel="epanechnikov", bw=7)

plot(d1, main="K.Epanechnikov")



par(mfrow=c(2,2))

d1<-density(faithful$waiting, kernel="gaussian")

plot(d1, main="K.Gaussian")

d1<-density(faithful$waiting, kernel="gaussian", bw=2)

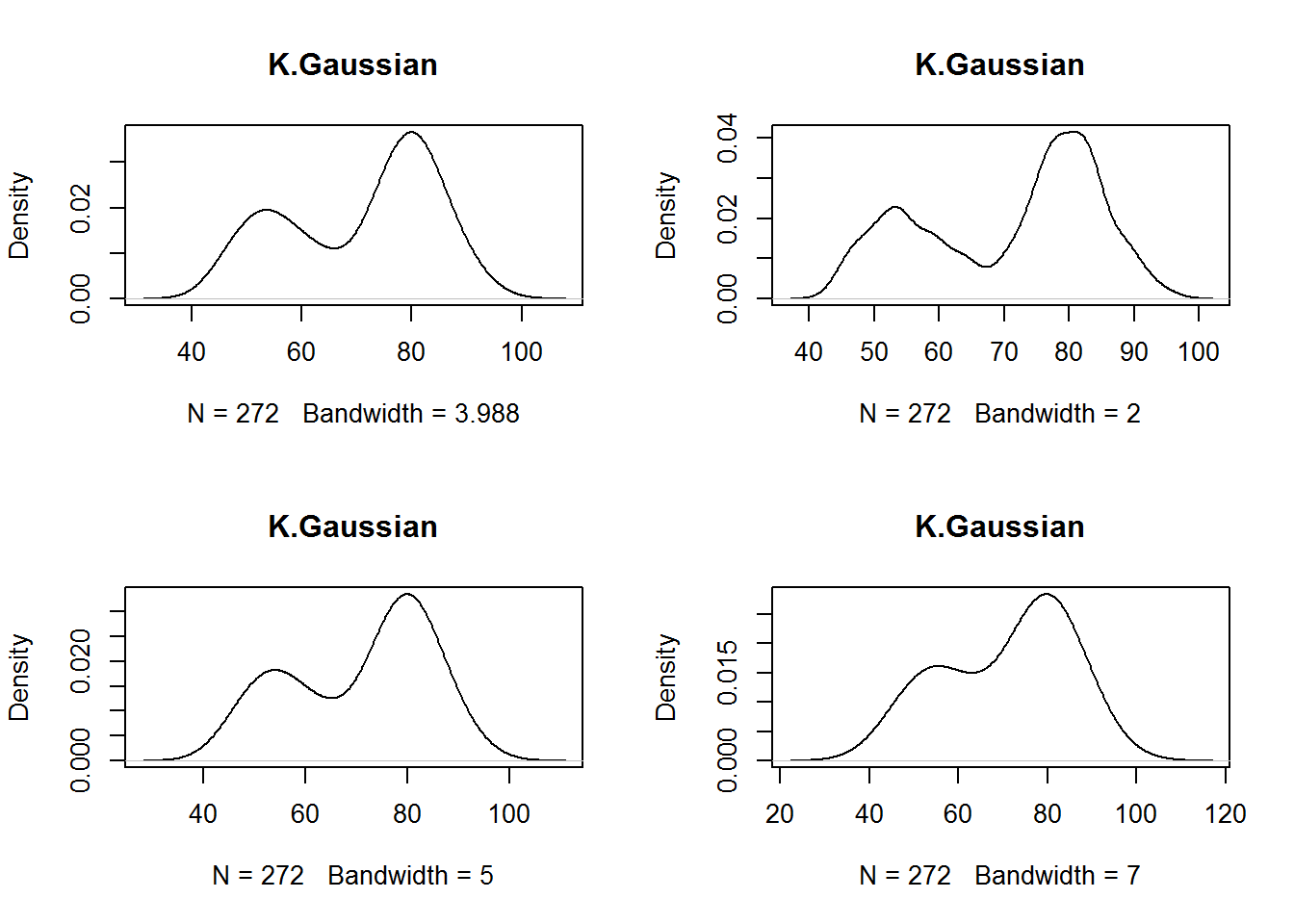
plot(d1, main="K.Gaussian")

d1<-density(faithful$waiting, kernel="gaussian", bw=5)

plot(d1, main="K.Gaussian")

d1<-density(faithful$waiting, kernel="gaussian", bw=7)

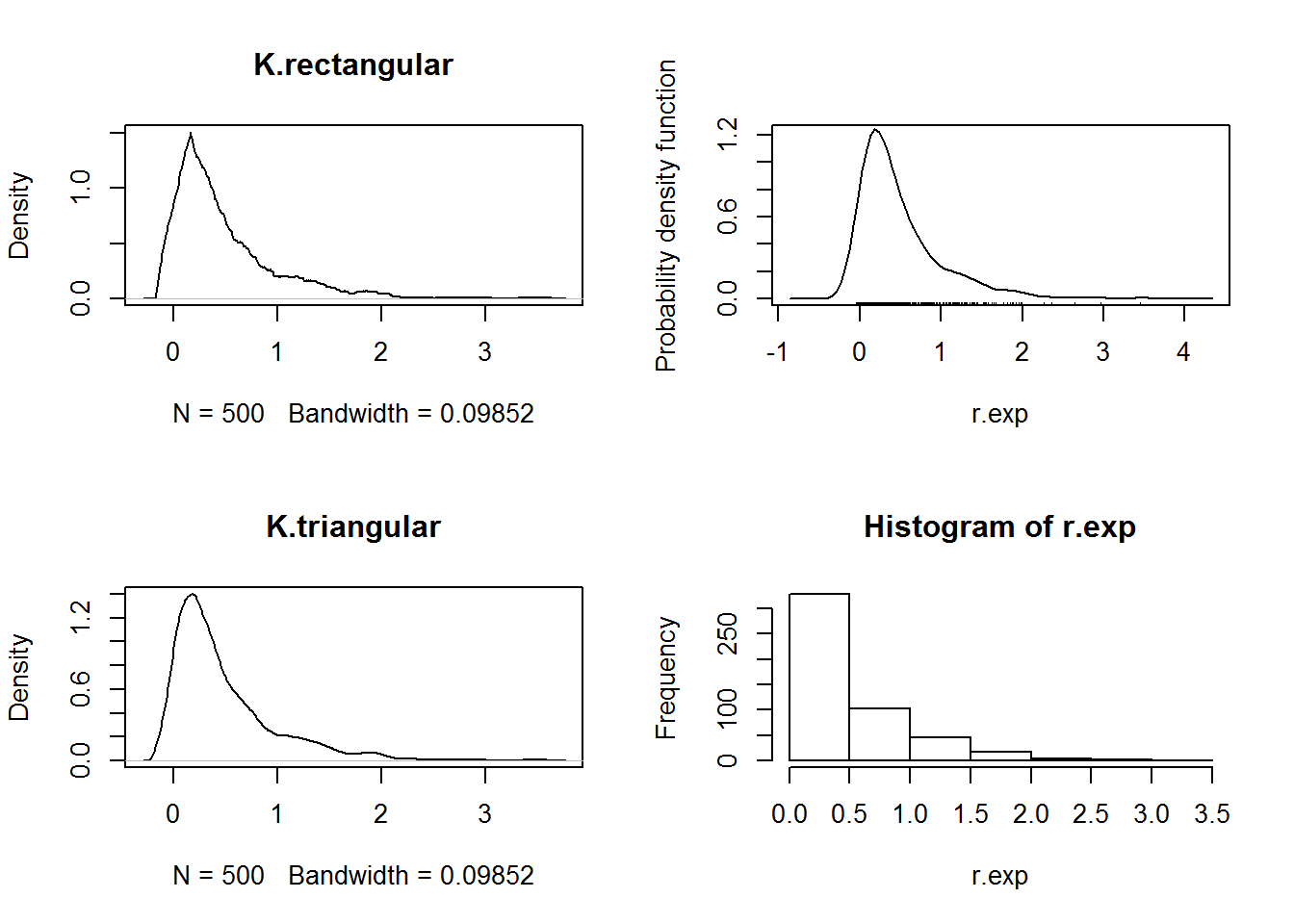
plot(d1, main="K.Gaussian")



(preguntar-se que passa amb l’apuntalament)

(Com triar el millor kernel? Dividir les dades en dos, estimar la densitat amb un i evaluar l’altre grup, Proposar m.bootstrap per veure que passa amb el baixi i amb la variabilitat)

(preguntar-se que passa amb una exponencial)



1. Expliqueu amb detall que ens mostra la següent sortida de la funció sm.density...

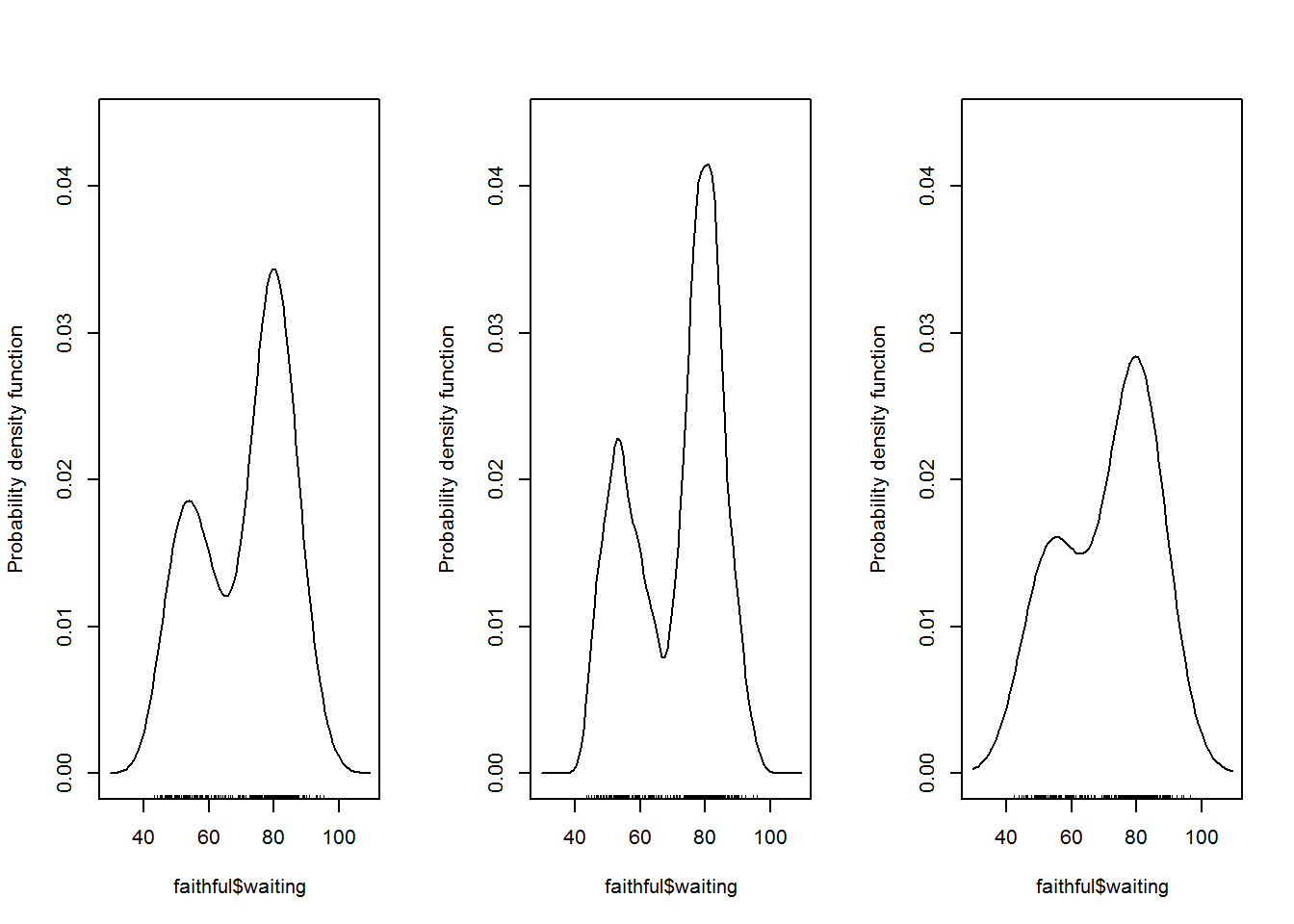
**require**(sm)

par(mfrow=c(1,3))

sm.density(faithful$waiting)

sm.density(faithful$waiting, h=2)

sm.density(faithful$waiting, h=7)

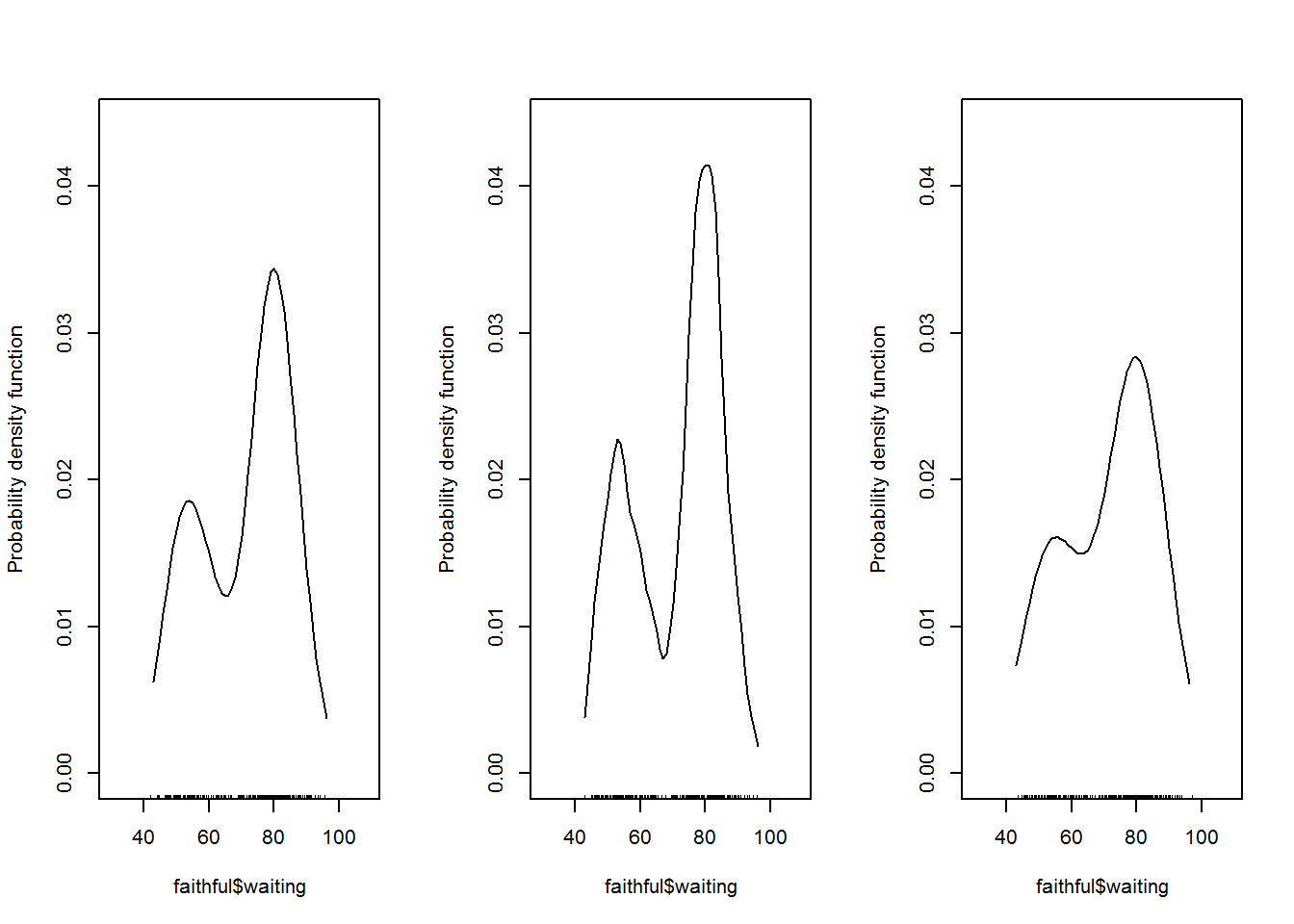


par(mfrow=c(1,3))

sm.density(faithful$waiting, eval.points=faithful$waiting[order(faithful$waiting)])

sm.density(faithful$waiting, h=2, eval.points=faithful$waiting[order(faithful$waiting)])

sm.density(faithful$waiting, h=7, eval.points=faithful$waiting[order(faithful$waiting)])



1. **Regressió Robusta i/o No paramètrica (Teoria)**
2. Expliqueu quins les semblances i diferències entre la regressió robusta i/o No paramètrica efectuada amb ajustos locals i la regressió robusta i/o No paramètrica efectuada amb restriccions.
3. Determineu quin dels tipus de regressió robusta i/o No paramètrica vistos a classe triaríeu en la següent situació;
   1. “es vol determinar si existeix relació entre dues variables i sabem que la relació no és monotònica”
   2. Volem ajustar un model no paramètric amb dues variables independents
   3. Volem determinar un grup reduït de variables independents a partir d’un gran nombre de variables.
4. Expliqueu ens quins cassos és més convenient utilitzar la regressió no paramètrica en comptes de la tau de Kendall. Poseu un exemple i expliqueu-lo amb claredat.
5. Quina és la diferència entre regressió robusta i regressió no paramètrica?
6. Que és un regressograma i quina és la seva utilitat?
7. **Regressió Robusta i/o No paramètrica (R)**
8. Trieu el llistat de R més convenient per la següent situació i comenteu per que

Es desconeix si existeix i de quin tipus seria la relació entre les variables temps d’espera (waiting) i durada de l’erupció (eruptions).

cor.test(faithful$waiting, faithful$eruptions, method="pearson", continuity = FALSE)

##

## Pearson's product-moment correlation

##

## data: faithful$waiting and faithful$eruptions

## t = 34.089, df = 270, p-value < 2.2e-16

## alternative hypothesis: true correlation is not equal to 0

## 95 percent confidence interval:

## 0.8756964 0.9210652

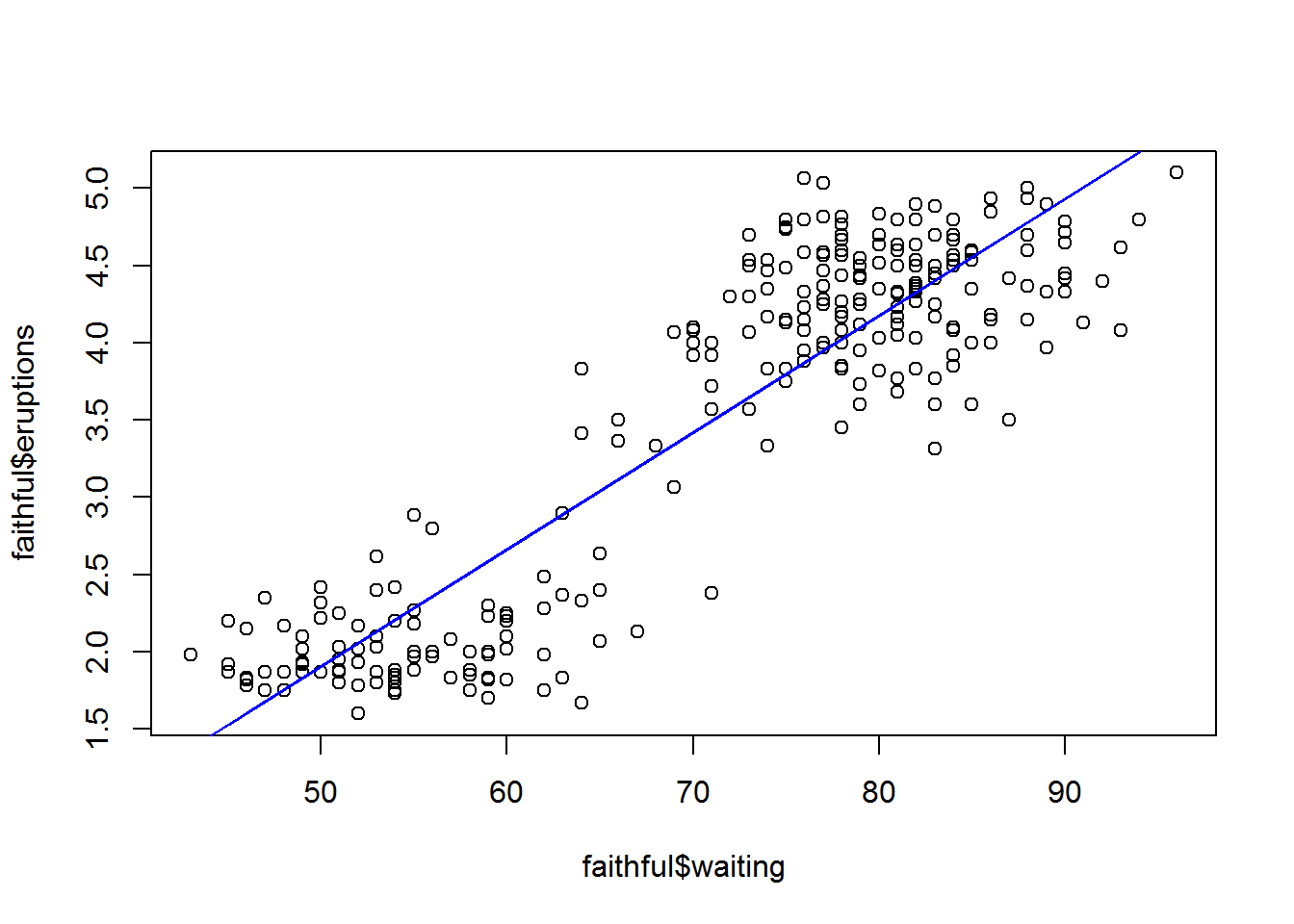
## sample estimates:

## cor

## 0.9008112

plot(faithful$waiting, faithful$eruptions)

lines(faithful$waiting,predict(lm(eruptions ~ waiting, data = faithful)),type="l",col=4)



cor.test(faithful$waiting, faithful$eruptions, method="kendall", continuity = FALSE)

##

## Kendall's rank correlation tau

##

## data: faithful$waiting and faithful$eruptions

## z = 13.902, p-value < 2.2e-16

## alternative hypothesis: true tau is not equal to 0

## sample estimates:

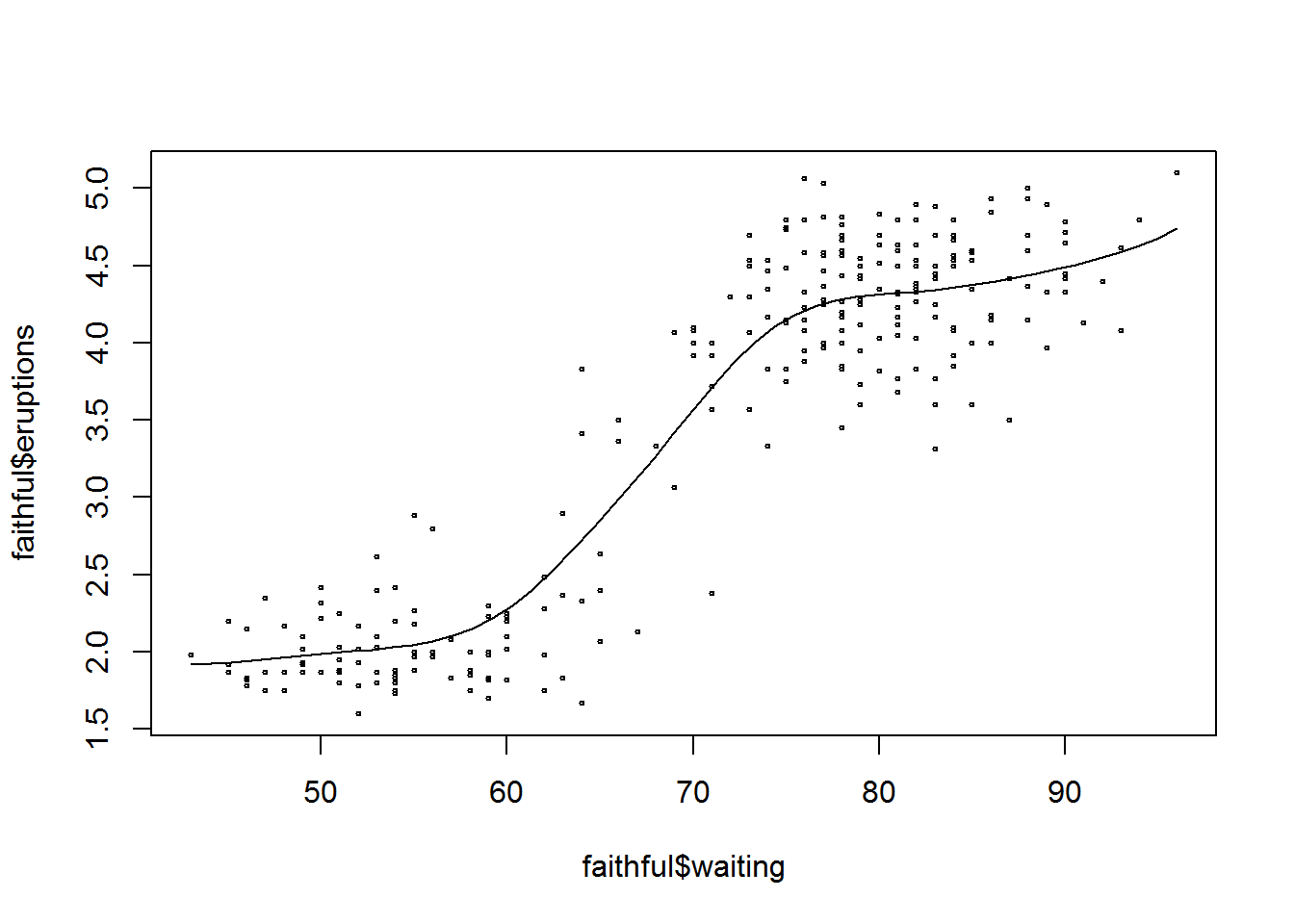
## tau

## 0.5747674

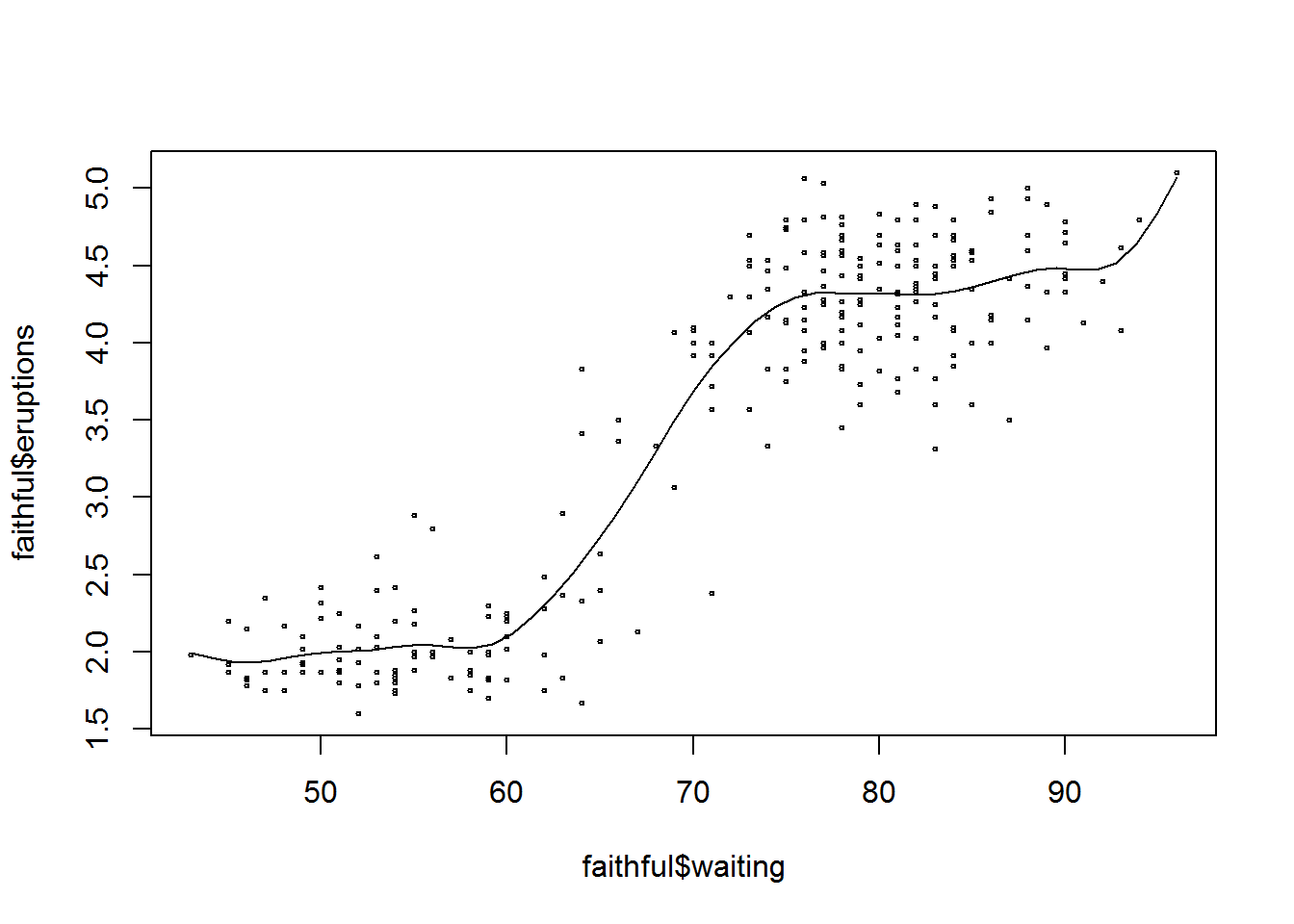
np1<-sm.regression(faithful$waiting, faithful$eruptions)

np1$h

## [1] 4.635649



np2<-sm.regression(faithful$waiting, faithful$eruptions, h=2)



1. Expliqueu com s’ha realitzat l’ajust obtingut pel següent codi i quines són les desavantatges d’aquest mètode?

nbins<-5

z <- cut(faithful$waiting,breaks=seq(min(faithful$waiting),

max(faithful$waiting),length=nbins+1), labels=1:nbins,include.lowest=TRUE)

ff.lm1 <- lm(eruptions[z==1] ~ waiting[z==1], data = faithful)

ff.lm2 <- lm(eruptions[z==2] ~ waiting[z==2], data = faithful)

ff.lm3 <- lm(eruptions[z==3] ~ waiting[z==3], data = faithful)

ff.lm4 <- lm(eruptions[z==4] ~ waiting[z==4], data = faithful)

ff.lm5 <- lm(eruptions[z==5] ~ waiting[z==5], data = faithful)

b1<-max(faithful$waiting[z==1])

b2<-max(faithful$waiting[z==2])

b3<-max(faithful$waiting[z==3])

b4<-max(faithful$waiting[z==4])

plot(faithful$waiting, faithful$eruptions)

*#Detectar els punts de tall*

abline(v=c(b1,b2,b3, b4), col=2)

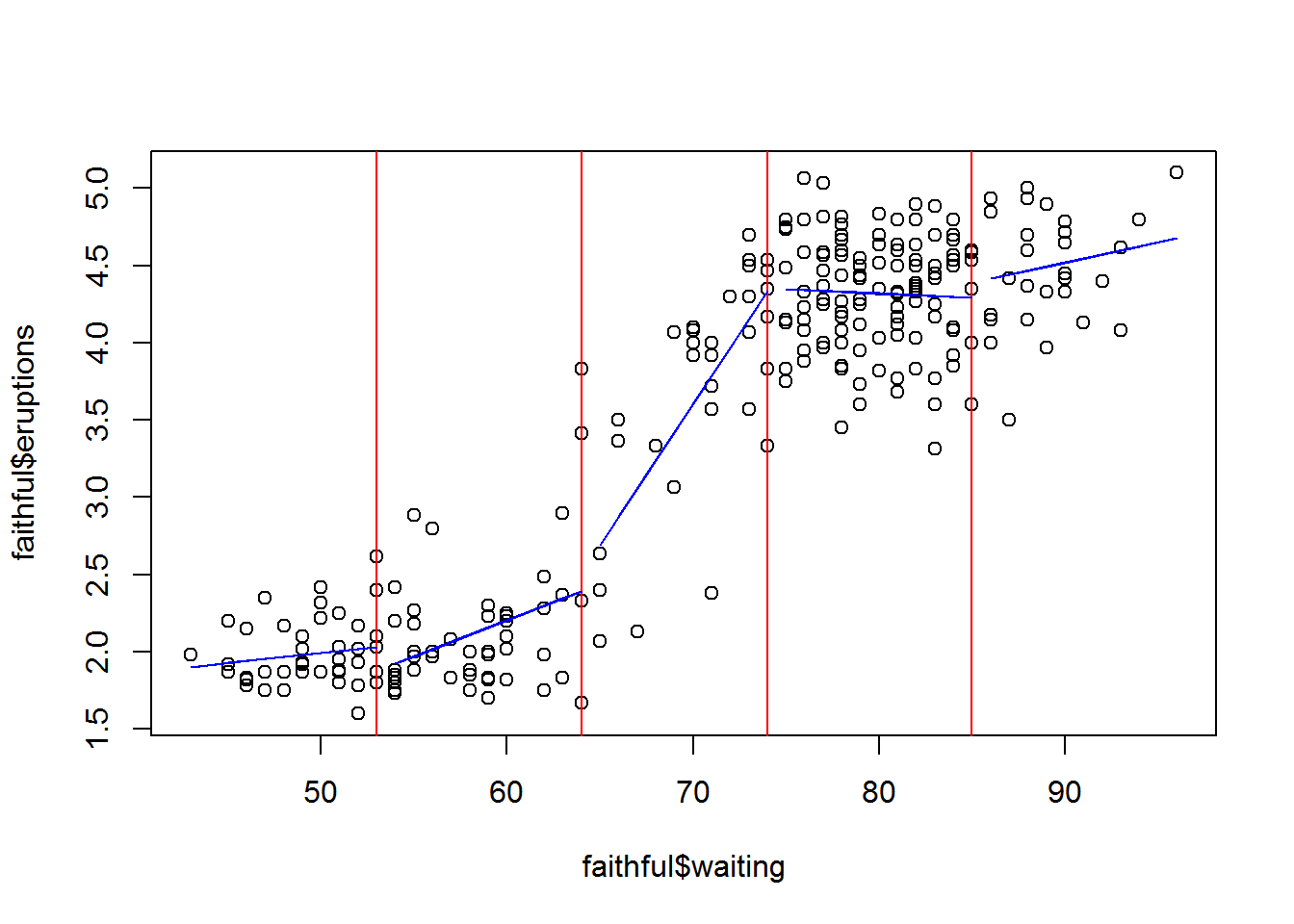
lines(faithful$waiting[z==1],predict(ff.lm1),type="l",col=4)

lines(faithful$waiting[z==2],predict(ff.lm2),type="l",col=4)

lines(faithful$waiting[z==3],predict(ff.lm3),type="l",col=4)

lines(faithful$waiting[z==4],predict(ff.lm4),type="l",col=4)

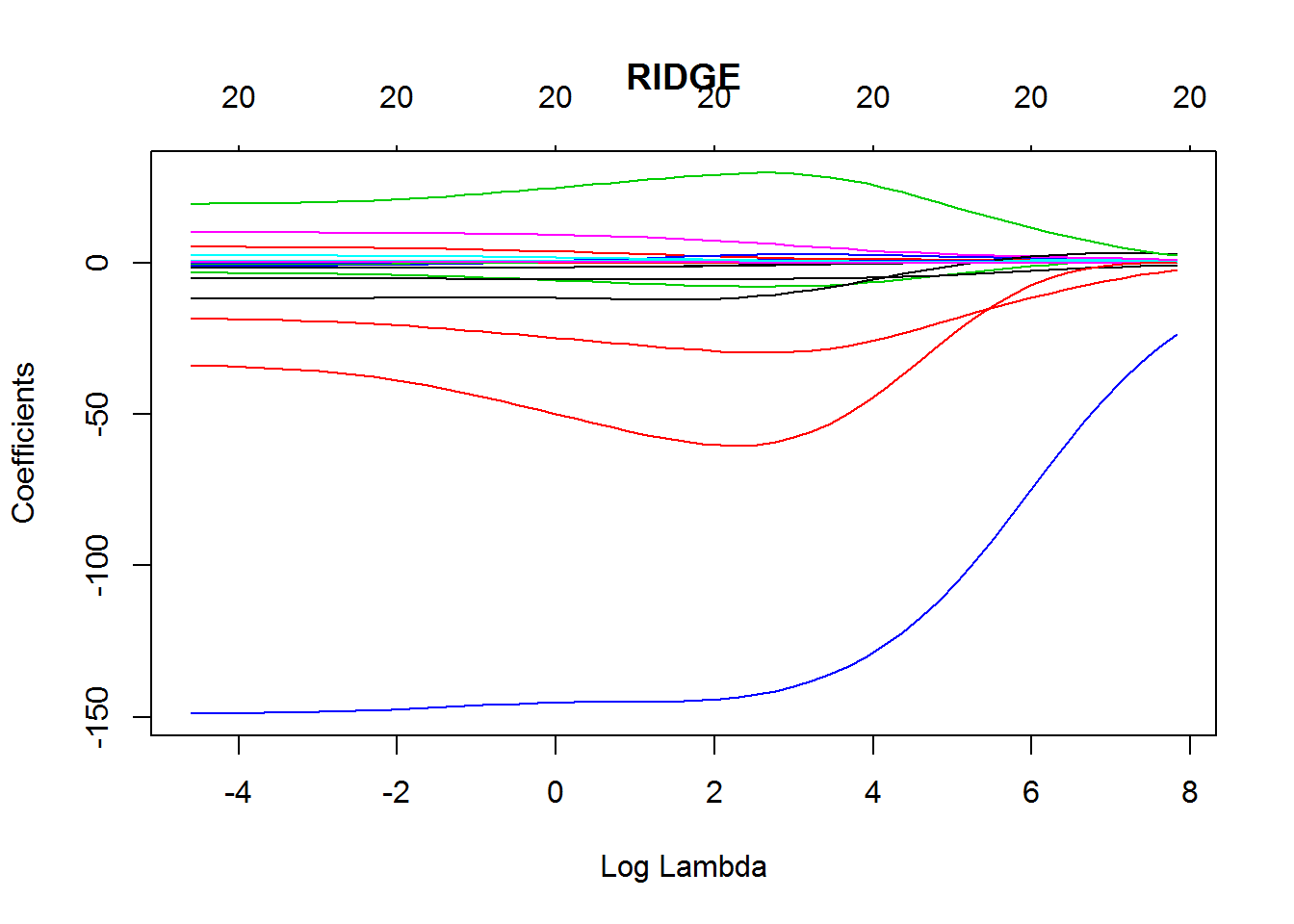
lines(faithful$waiting[z==5],predict(ff.lm5),type="l",col=4)

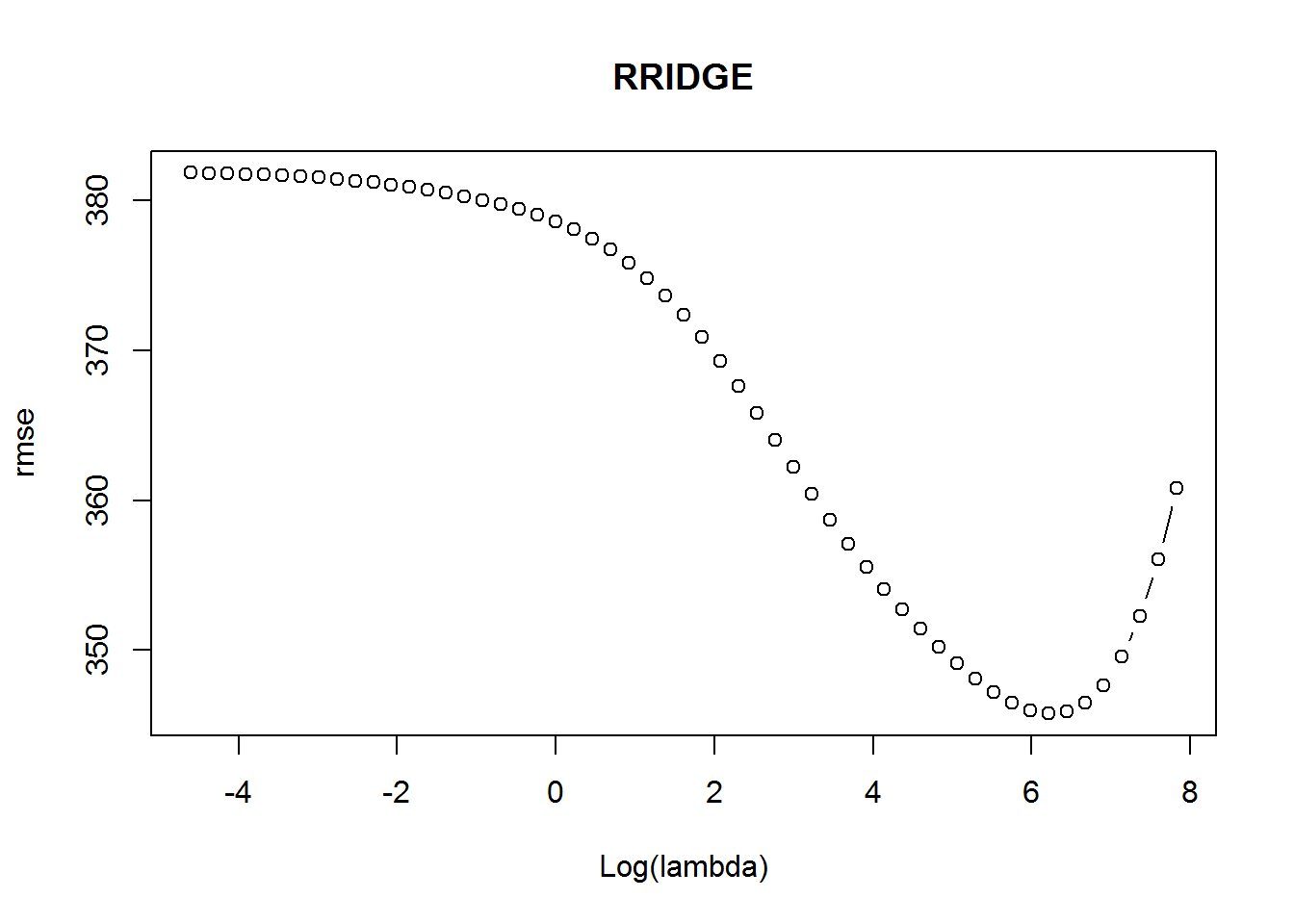


1. Donat el codi següent, comenteu que s’ha realitzat i discutiu quina seria la conclusió a la que arribaríeu.

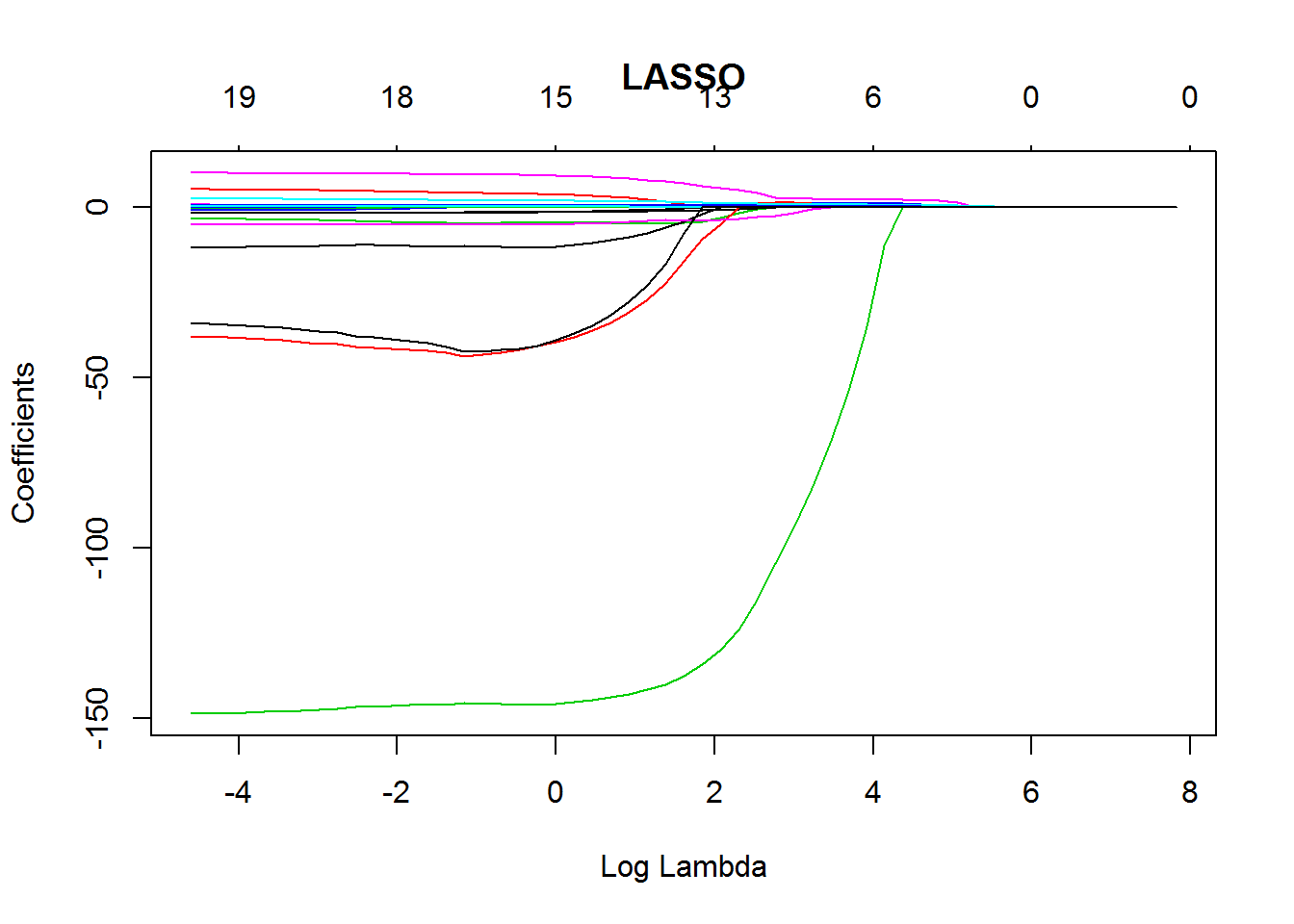
lambdas = 10^seq(-2,3.4,0.1)

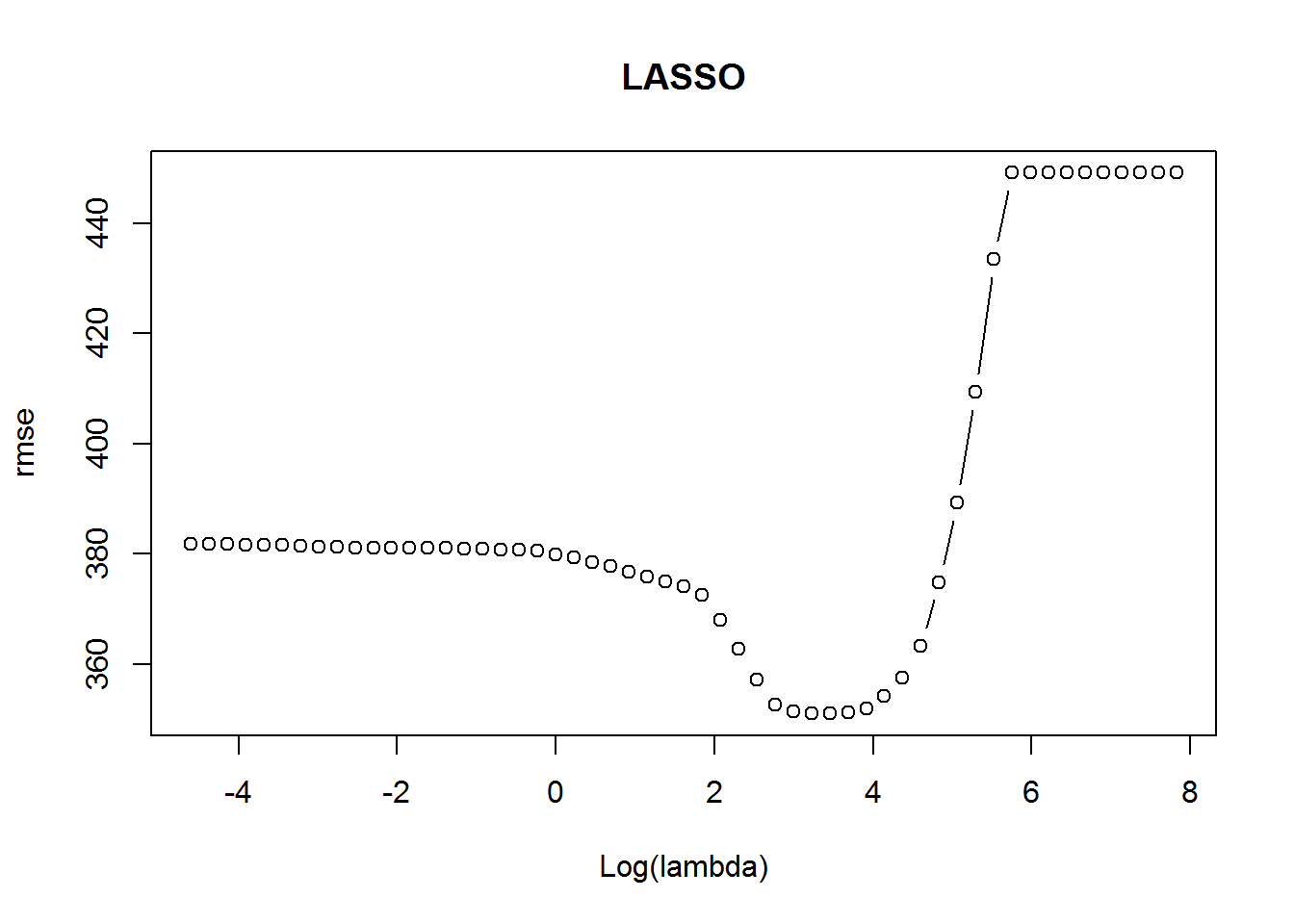
fm.lasso = glmnet(X.train,Y.train, alpha = 0, lambda = lambdas, thresh = 1e-12)





fm.lasso = glmnet(X.train,Y.train, alpha = 1, lambda = lambdas, thresh = 1e-12)





Comproveu que passaria si les dades estiguessin estanderitzades