#Set the seed

set.seed(4428967)

#Set the working dir

setwd("F:/vigupta/OneDrive/Learning/DataScience/Statistics Texas A&M University/689/Assignment/Assignment12")

#libraries

library(HRW)

library(lattice)

library(nlme)

#data on repeated measures on the pigs

pigWeights = read.csv("pigWeights(1).csv")

attach(pigWeights)

#explore , we will treat num weeks as nominal

head(pigWeights , 10)

###

###Answer to question 1.a

###

length(unique(pigWeights$id.num))

#48

###

###Answer to question 1.b. Please see the summary toward later stage of the code

###

#The ids are sequential

###

###Hard setup of the random structure and the splines for the fixed effects

###since gamm in mgcv cannot do it

###

#Number of obs.

numObs = length(weight)

# Number of subjects

numGrp = length(unique(id.num))

# Number of knots at the population level

numIntKnotsGbl = 20

# O'Sullivan Basis Functions at the population level

intKnotsGbl = quantile(unique(num.weeks),

seq(0,1,length=numIntKnotsGbl+2))[-c(1,numIntKnotsGbl+2)]

range.weeks = c(min(num.weeks)-0.01, max(num.weeks)+0.01)

Zgbl = ZOSull(num.weeks,range.x=range.weeks,intKnots=intKnotsGbl)

# Number of knots at the subject level. Should be less than the

# number of knots at the population level

numIntKnotsGrp = 3

# Basis functions for O'Sullivan splines at the

# individual level

intKnotsGrp = quantile(unique(num.weeks),

seq(0,1,length=numIntKnotsGrp+2))[-c(1,numIntKnotsGrp+2)]

Zgrp = ZOSull(num.weeks,range.x=range.weeks,intKnots=intKnotsGrp)

#####################################################################

# Set up the random effects structure and call lme

#

# See Lecture 18 for the explanation

#####################################################################

#

dummyId = factor(rep(1,numObs))

Zblock = list(dummyId = pdIdent( ~ -1 + Zgbl),

id.num = pdSymm( ~ num.weeks),

id.num = pdIdent( ~ -1 + Zgrp))

pigWeightGD = groupedData(weight ~ num.weeks|rep(1,length = numObs),

data = data.frame(weight,num.weeks,Zgbl,Zgrp,id.num))

fit = lme(weight ~ num.weeks,data = pigWeightGD,random = Zblock)

##Ignore the error in the fit, possibly due to more knots than measurements

summary(fit)

####Linear mixed-effects model fit by REML

#### Data: pigWeightGD

#### AIC BIC logLik

#### 1646.664 1679.174 -815.3321

####

####Random effects:

#### Formula: ~-1 + Zgbl | dummyId

#### Structure: Multiple of an Identity

#### Zgbl1 Zgbl2 Zgbl3 Zgbl4 Zgbl5 Zgbl6 Zgbl7 Zgbl8 Zgbl9 Zgbl10 Zgbl11

####StdDev: 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533

#### Zgbl12 Zgbl13 Zgbl14 Zgbl15 Zgbl16 Zgbl17 Zgbl18 Zgbl19 Zgbl20 Zgbl21 Zgbl22

####StdDev: 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533

####

#### Formula: ~num.weeks | id.num %in% dummyId

#### Structure: General positive-definite

#### StdDev Corr

####(Intercept) 2.6885425 (Intr)

####num.weeks 0.6291173 -0.098

####

#### Formula: ~-1 + Zgrp | id.num %in% id.num %in% dummyId

#### Structure: Multiple of an Identity

#### Zgrp1 Zgrp2 Zgrp3 Zgrp4 Zgrp5 Zgrp6 Zgrp7 Zgrp8 Zgrp9 Zgrp10 Zgrp11

####StdDev: 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216

#### Zgrp12 Residual

####StdDev: 0.6421216 0.8354407

####

####Fixed effects: weight ~ num.weeks

#### Value Std.Error DF t-value p-value

####(Intercept) 19.358295 0.3999824 383 48.39787 0

####num.weeks 6.211238 0.0924196 383 67.20697 0

#### Correlation:

#### (Intr)

####num.weeks -0.133

####

####Standardized Within-Group Residuals:

#### Min Q1 Med Q3 Max

####-3.070195309 -0.462303093 -0.002530952 0.433184383 2.479136098

####

####Number of Observations: 432

####Number of Groups:

#### dummyId id.num %in% dummyId id.num.1 %in% id.num %in% dummyId

#### 1 48 48

sigma\_grp = 0.8354407 # from the fit.

sigma\_gbl = 0.9017533

#####################################################################

# Now we are going to get BLUPs and plot functions,

# This is for the popn level estimates and not at individual levels

#####################################################################

# Size of grid on which to do the plotting

ng = 101

# Get the grid of ages

weekg = seq(range.weeks[1],range.weeks[2],length = ng)

# Form the design matrix for the linear part, which is a column

# of ones plus the grid of ages

Xg = cbind(rep(1,ng),weekg)

# Spline terms for the overall fit

Zgblg = ZOSull(weekg,range.x = range.weeks,

intKnots = intKnotsGbl)

# Spline terms for the individual fits

Zgrpg = ZOSull(weekg,range.x = range.weeks,

intKnots = intKnotsGrp)

# Get betahat, the intercept and the slope

betaHat = as.vector(fit$coef$fixed)

# Get uHat, the estimated spline coefficients for the

# overall fit

uHat = as.vector(fit$coef$random[[1]])

# Form the overall fit

fHatg = as.vector(Xg%\*%betaHat + Zgblg%\*%uHat)

###

###Answer to question 2

###

#####################################################################

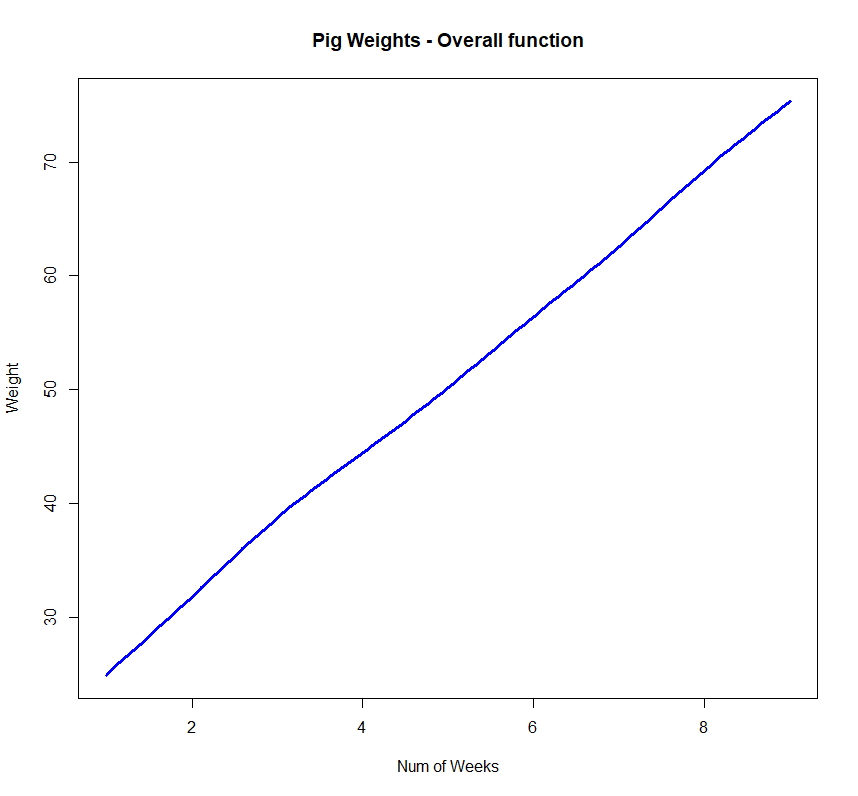
# Plot the overall function. BLUP estimates

#####################################################################

plot(weekg,fHatg,type='l',lwd=3,col="blue",

xlab="Num of Weeks",ylab="Weight",

main="Pig Weights - Overall function" )



###

###Answer to question 3

###

#####################################################################

# Plot the subject specific function. BLUP estimates. Note: This may

# not be required always. We are plotting only 28 pigs individualy over weeks

#####################################################################

# Now get the subject-specific estimated curves

curvEsts = vector("list",numGrp)

for (i in 1:numGrp)

{

# The subject-specific terms for the slope

# and intercept

uLinHati = as.vector(fit$coef$random[[2]][i,])

# The subject-specific terms for the spline coefficients

uSplHati = as.vector(fit$coef$random[[3]][i,])

# The individual function estimates

ghati = Xg%\*%uLinHati + Zgrpg%\*%uSplHati

curvEsts[[i]] = fHatg + ghati

}

###Lattice plot

pigWeightFit\_Subject = xyplot(weight ~ num.weeks|id.num[id.num %in% 1:28],groups = id.num[id.num %in% 1:28],

data = pigWeights,

strip = FALSE,scales = list(cex = 1.25),

xlab = list("Num of weeks",cex = 1.5),

ylab = list("Weight",cex = 1.5),

as.table = TRUE,#layout = c(4,7),

panel = function(x,y,subscripts,groups)

{

panel.grid()

adolNum = id.num[id.num %in% 1:28][subscripts][1]

panel.superpose(x,y,subscripts,groups,

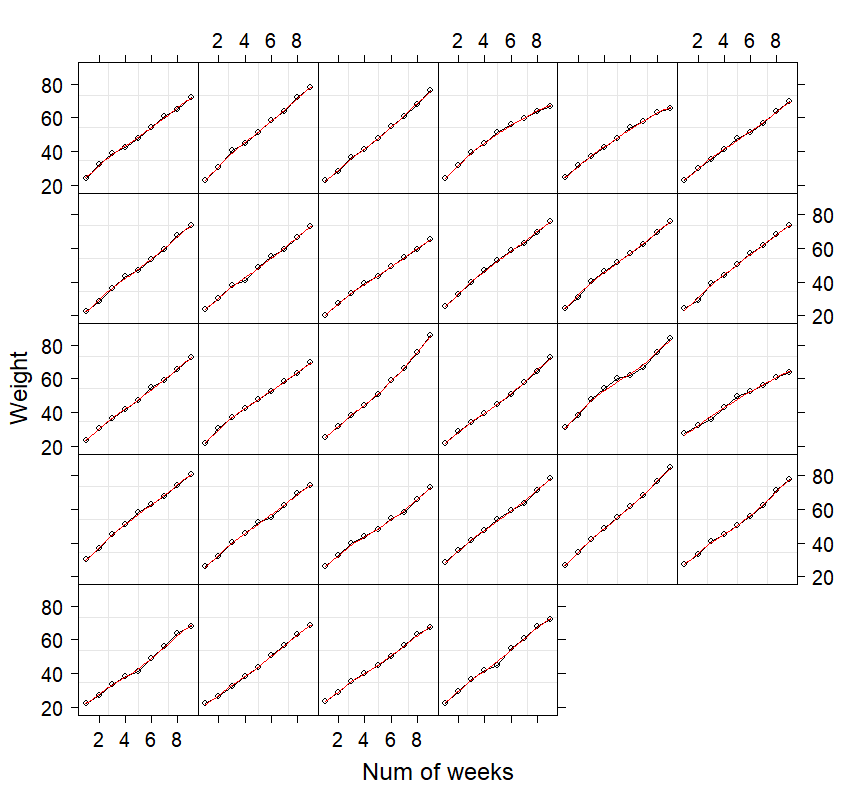
col = "black",type = "b")

panel.xyplot(weekg,curvEsts[[adolNum]],

col = "red",type = "l")

})

plot(pigWeightFit\_Subject)



###

###Answer to question 4. We do find the between subject variability more in weeks 6-9 than in weeks 1-5

###sigma\_grp1 = 0.6897996 , sigma\_grp2 = 0.7374592

###

###Analysis to find siga\_grp for period/weeks 1-5. After fitting we get the estimates as below

###sigma\_grp1 = 0.6897996 # from the fit.

###sigma\_gbl1 = 1.167525

###

detach(pigWeights)

pigWeights = read.csv("pigWeights(1).csv")

pigWeights = pigWeights[ pigWeights$num.weeks %in% 1:5 , ]

attach(pigWeights)

#Number of obs.

numObs = length(weight)

# Number of subjects

numGrp = length(unique(id.num))

# Number of knots at the population level

numIntKnotsGbl = 20

# O'Sullivan Basis Functions at the population level

intKnotsGbl = quantile(unique(num.weeks),

seq(0,1,length=numIntKnotsGbl+2))[-c(1,numIntKnotsGbl+2)]

range.weeks = c(min(num.weeks)-0.01, max(num.weeks)+0.01)

Zgbl = ZOSull(num.weeks,range.x=range.weeks,intKnots=intKnotsGbl)

# Number of knots at the subject level. Should be less than the

# number of knots at the population level

numIntKnotsGrp = 3

# Basis functions for O'Sullivan splines at the

# individual level

intKnotsGrp = quantile(unique(num.weeks),

seq(0,1,length=numIntKnotsGrp+2))[-c(1,numIntKnotsGrp+2)]

Zgrp = ZOSull(num.weeks,range.x=range.weeks,intKnots=intKnotsGrp)

#####################################################################

# Set up the random effects structure and call lme

#

# See Lecture 18 for the explanation

#####################################################################

#

dummyId = factor(rep(1,numObs))

Zblock = list(dummyId = pdIdent( ~ -1 + Zgbl),

id.num = pdSymm( ~ num.weeks),

id.num = pdIdent( ~ -1 + Zgrp))

pigWeightGD = groupedData(weight ~ num.weeks|rep(1,length = numObs),

data = data.frame(weight,num.weeks,Zgbl,Zgrp,id.num))

fit = lme(weight ~ num.weeks,data = pigWeightGD,random = Zblock)

##Ignore the error in the fit, possibly due to more knots than measurements

summary(fit)

###Linear mixed-effects model fit by REML

### Data: pigWeightGD

### AIC BIC logLik

### 876.0367 903.8148 -430.0183

###

###Random effects:

### Formula: ~-1 + Zgbl | dummyId

### Structure: Multiple of an Identity

### Zgbl1 Zgbl2 Zgbl3 Zgbl4 Zgbl5 Zgbl6 Zgbl7 Zgbl8 Zgbl9 Zgbl10 Zgbl11 Zgbl12

###StdDev: 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525

### Zgbl13 Zgbl14 Zgbl15 Zgbl16 Zgbl17 Zgbl18 Zgbl19 Zgbl20 Zgbl21 Zgbl22

###StdDev: 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525 1.167525

###

### Formula: ~num.weeks | id.num %in% dummyId

### Structure: General positive-definite

### StdDev Corr

###(Intercept) 2.1935981 (Intr)

###num.weeks 0.7343408 0.09

###

### Formula: ~-1 + Zgrp | id.num %in% id.num %in% dummyId

### Structure: Multiple of an Identity

### Zgrp1 Zgrp2 Zgrp3 Zgrp4 Zgrp5 Residual

###StdDev: 0.5820647 0.5820647 0.5820647 0.5820647 0.5820647 0.6897996

###

###Fixed effects: weight ~ num.weeks

### Value Std.Error DF t-value p-value

###(Intercept) 19.227686 0.3441630 191 55.86797 0

###num.weeks 6.291623 0.1137349 191 55.31829 0

### Correlation:

### (Intr)

###num.weeks -0.053

###

###Standardized Within-Group Residuals:

### Min Q1 Med Q3 Max

###-2.49495716 -0.42130791 0.03167879 0.47271404 2.03886898

###

###Number of Observations: 240

###Number of Groups:

### dummyId id.num %in% dummyId id.num.1 %in% id.num %in% dummyId

### 1 48 48

sigma\_grp1 = 0.6897996 # from the fit.

sigma\_gbl1 = 1.167525

###

###Analysis to find siga\_grp for period/weeks 6-9. After fitting we get the estimates as below

###sigma\_grp2 = 0.7374592 # from the fit.

###sigma\_gbl2 = 1.089808

###

detach(pigWeights)

pigWeights = read.csv("pigWeights(1).csv")

pigWeights = pigWeights[ pigWeights$num.weeks %in% 6:9 , ]

attach(pigWeights)

#Number of obs.

numObs = length(weight)

# Number of subjects

numGrp = length(unique(id.num))

# Number of knots at the population level

numIntKnotsGbl = 20

# O'Sullivan Basis Functions at the population level

intKnotsGbl = quantile(unique(num.weeks),

seq(0,1,length=numIntKnotsGbl+2))[-c(1,numIntKnotsGbl+2)]

range.weeks = c(min(num.weeks)-0.01, max(num.weeks)+0.01)

Zgbl = ZOSull(num.weeks,range.x=range.weeks,intKnots=intKnotsGbl)

# Number of knots at the subject level. Should be less than the

# number of knots at the population level

numIntKnotsGrp = 2

# Basis functions for O'Sullivan splines at the

# individual level

intKnotsGrp = quantile(unique(num.weeks),

seq(0,1,length=numIntKnotsGrp+2))[-c(1,numIntKnotsGrp+2)]

Zgrp = ZOSull(num.weeks,range.x=range.weeks,intKnots=intKnotsGrp)

#####################################################################

# Set up the random effects structure and call lme

#

# See Lecture 18 for the explanation

#####################################################################

#

dummyId = factor(rep(1,numObs))

Zblock = list(dummyId = pdIdent( ~ -1 + Zgbl),

id.num = pdSymm( ~ num.weeks),

id.num = pdIdent( ~ -1 + Zgrp))

pigWeightGD = groupedData(weight ~ num.weeks|rep(1,length = numObs),

data = data.frame(weight,num.weeks,Zgbl,Zgrp,id.num))

fit = lme(weight ~ num.weeks,data = pigWeightGD,random = Zblock)

##Ignore the error in the fit, possibly due to more knots than measurements

summary(fit)

###Linear mixed-effects model fit by REML

### Data: pigWeightGD

### AIC BIC logLik

### 827.4544 853.4305 -405.7272

###

###Random effects:

### Formula: ~-1 + Zgbl | dummyId

### Structure: Multiple of an Identity

### Zgbl1 Zgbl2 Zgbl3 Zgbl4 Zgbl5 Zgbl6 Zgbl7 Zgbl8 Zgbl9 Zgbl10 Zgbl11 Zgbl12

###StdDev: 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808

### Zgbl13 Zgbl14 Zgbl15 Zgbl16 Zgbl17 Zgbl18 Zgbl19 Zgbl20 Zgbl21 Zgbl22

###StdDev: 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808 1.089808

###

### Formula: ~num.weeks | id.num %in% dummyId

### Structure: General positive-definite

### StdDev Corr

###(Intercept) 5.9545297 (Intr)

###num.weeks 0.9995404 -0.729

###

### Formula: ~-1 + Zgrp | id.num %in% id.num %in% dummyId

### Structure: Multiple of an Identity

### Zgrp1 Zgrp2 Zgrp3 Zgrp4 Residual

###StdDev: 1.051654 1.051654 1.051654 1.051654 0.7374592

###

###Fixed effects: weight ~ num.weeks

### Value Std.Error DF t-value p-value

###(Intercept) 17.979973 0.9823698 143 18.30265 0

###num.weeks 6.384449 0.1573706 143 40.56951 0

### Correlation:

### (Intr)

###num.weeks -0.776

###

###Standardized Within-Group Residuals:

### Min Q1 Med Q3 Max

###-1.75945430 -0.42543698 0.00730456 0.39604790 2.28977088

###

###Number of Observations: 192

###Number of Groups:

### dummyId id.num %in% dummyId id.num.1 %in% id.num %in% dummyId

### 1 48 48

sigma\_grp2 = 0.7374592 # from the fit.

sigma\_gbl2 = 1.089808