Assignment 12; STAT 689

Philip Anderson; panders2@tamu.edu

4/10/2018

```
# preliminaries
rm(list=ls())
library("HRW")
library("mgcv")
library("nlme")
library("lattice")
library("tidyverse")
```

```
# bring in the data
pigs <- read.csv('/Users/panders2/Documents/schools/tamu/stat_689/homework/semiparametric-regression/hw
str(pigs)
```

```
## 'data.frame':    432 obs. of  3 variables:
##  $ id.num   : int  1 1 1 1 1 1 1 1 1 2 ...
##  $ num.weeks: int  1 2 3 4 5 6 7 8 9 1 ...
##  $ weight   : num  24 32 39 42.5 48 54.5 61 65 72 22.5 ...
```

Question 1

1A

How many pigs are in the model?

```
cat('Distinct pigs in the data given by: ', length(unique(pigs$id.num)))
```

```
## Distinct pigs in the data given by:  48
```

1B

Fit the random function model and display your code.
I am going to model pig weight as a function of the number of weeks since measurement on the pigs began.

```
# extract the important variables into individual objects
id_num_a <- pigs$id.num
num_weeks_a <- pigs$num.weeks
pig_weight_a <- pigs$weight
```

Now, we need to set up the design matrices for the splines at the population (global) level and individual (group) level. Note that the individual level will have fewer knots than the population level. We will follow this with a random effects structure, generated below. There is a lot of code here that will be reused later, so I am going to package this up into a function - note that it won't be generalizable beyond this specific situation.

```
random_func_mod <- function(id_num, num_weeks, pig_weight) {

      # a lot of these objects get used down the road, so I am going to use global assignment
```

```r
      # number of records
      numObs <<- length(id_num)

      # number of subjects
      numGrp <<- length(unique(id_num))

      # population (Gbl) work
      # knots
      numIntKnotsGbl <<- 20
      # population O-Sull Basis Functions
      intKnotsGbl <<- quantile(unique(num_weeks)
                              , seq(0, 1, length=numIntKnotsGbl+2)
                              )[-c(1, numIntKnotsGbl+2)]
      range.num_weeks <<- c(min(num_weeks)-0.01, max(num_weeks)+0.01)

      Zgbl <<- HRW::ZOSull(num_weeks, range.x=range.num_weeks, intKnots=intKnotsGbl)

      # subject-level (Grp) work
      numIntKnotsGrp <<- 10
      intKnotsGrp <<- quantile(unique(num_weeks)
                              , seq(0, 1, length=numIntKnotsGrp+2)
                              )[-c(1, numIntKnotsGrp+2)]
      Zgrp <<- HRW::ZOSull(num_weeks, range.x=range.num_weeks, intKnots=intKnotsGrp)


      #Now, set up the random effects structure.
      dummyId <<- factor(rep(1, numObs))
      Zblock <<- list(
                   dummyId = pdIdent( ~ -1 + Zgbl)
                   , id_num = pdSymm(~ num_weeks)
                   , id_num = pdIdent(~ -1 + Zgrp)
                   )
      gd <<- groupedData(pig_weight ~ num_weeks|rep(1, length=numObs)
                        , data=data.frame(pig_weight, num_weeks, Zgbl, Zgrp, id_num)
                        )
      fit <<- lme(pig_weight ~ num_weeks, data=gd, random=Zblock)

      return(fit)

}
```

```r
fit <- random_func_mod(id_num=id_num_a, num_weeks=num_weeks_a, pig_weight=pig_weight_a)
```

```
## Warning in lme.formula(pig_weight ~ num_weeks, data = gd, random = Zblock):
## fewer observations than random effects in all level 3 groups
```

Display the summary.

```r
summary(fit)
```

```
## Linear mixed-effects model fit by REML
##  Data: gd
##        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
## StdDev: 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533
##             Zgbl7     Zgbl8     Zgbl9    Zgbl10    Zgbl11    Zgbl12
## StdDev: 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533
##            Zgbl13    Zgbl14    Zgbl15    Zgbl16    Zgbl17    Zgbl18
## StdDev: 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533 0.9017533
##            Zgbl19    Zgbl20    Zgbl21    Zgbl22
## StdDev: 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
## StdDev: 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216
##             Zgrp7     Zgrp8     Zgrp9    Zgrp10    Zgrp11    Zgrp12
## StdDev: 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216 0.6421216
##          Residual
## StdDev: 0.8354407
##
## Fixed effects: pig_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  0.433184382  2.479136098
##
## Number of Observations: 432
## Number of Groups:
##                           dummyId                id_num %in% dummyId
##                                 1                                 48
## id_num.1 %in% id_num %in% dummyId
##                                48
```

Question 2

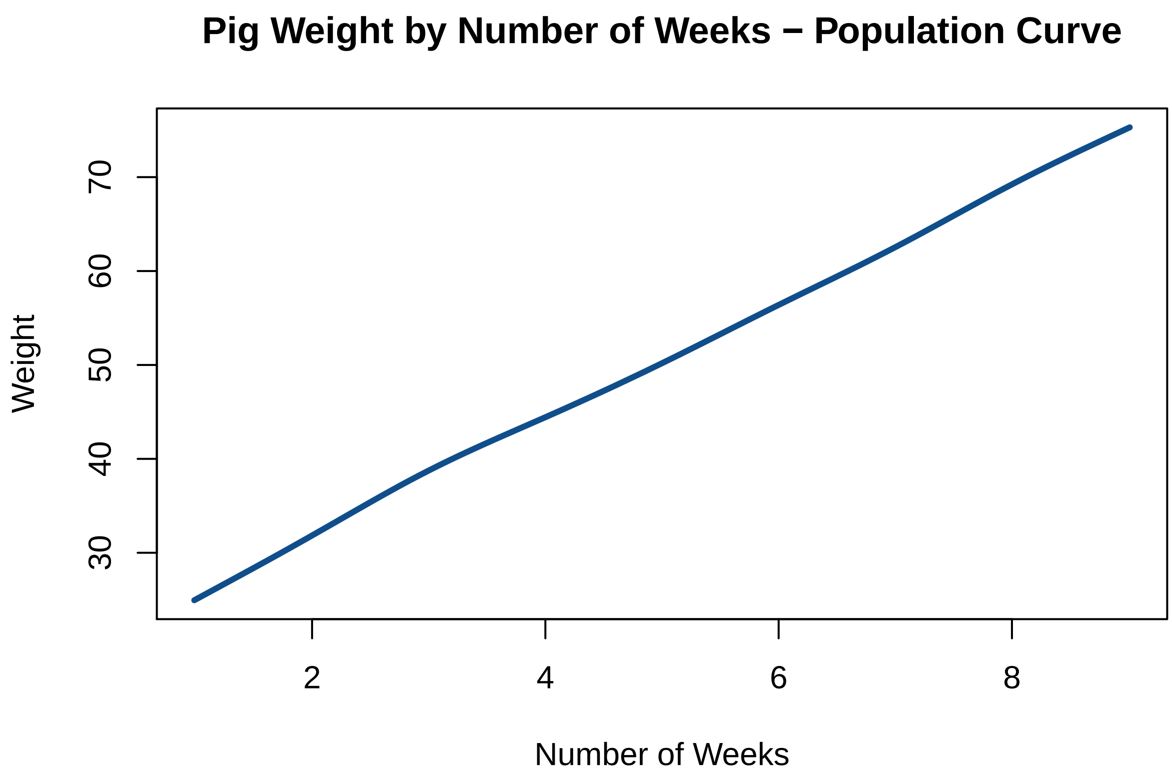
Plot the population-level BLUP estimates. First, we need to work through a number of preliminaries.

```
# number of grid points
ng <- 101
# grid for num_weeks
```

```r
num_weeks_g <- seq(range.num_weeks[1], range.num_weeks[2], length=ng)
# design matrix for linear component; col of 1's plus num_weeks grid
Xg <- cbind(rep(1, ng), num_weeks_g)

# spline terms - overall fit
Zgblg <- HRW::ZOSull(num_weeks_g, range.x=range.num_weeks, intKnots=intKnotsGbl)
# spline terms for individual fits
Zgrpg <- HRW::ZOSull(num_weeks_g, range.x=range.num_weeks, intKnots=intKnotsGrp)

# grab betaHat, the model intercept, and the slope from our model objet
betaHat <- as.vector(fit$coefficients$fixed)
# grab uHat, along with the estimated spline coef for overall fit
uHat <- as.vector(fit$coefficients$random[[1]])
# form the overall fit
fHatg <- as.vector(Xg %*% betaHat + Zgblg %*% uHat)
# subject-specific estimated curves
curvEsts <- vector("list", numGrp)

for (i in 1:numGrp)
  {
  # subject-specific slope + intercept
  uLinHati <- as.vector(fit$coefficients$random[[2]][i, ])
  # subject-specific terms for splines
  uSplHati <- as.vector(fit$coefficients$random[[3]][i, ])
  # individual function estimates
  ghati <- Xg %*% uLinHati + Zgrpg %*% uSplHati
  curvEsts[[i]] <- fHatg + ghati
  }
```

Now do the population-level plot.

```r
plot(num_weeks_g, fHatg, type="l", col="dodgerblue4", lwd=3
     , xlab="Number of Weeks"
     , ylab="Weight"
     , main="Pig Weight by Number of Weeks - Population Curve"
     )
```



Overall, this plot looks very linear.

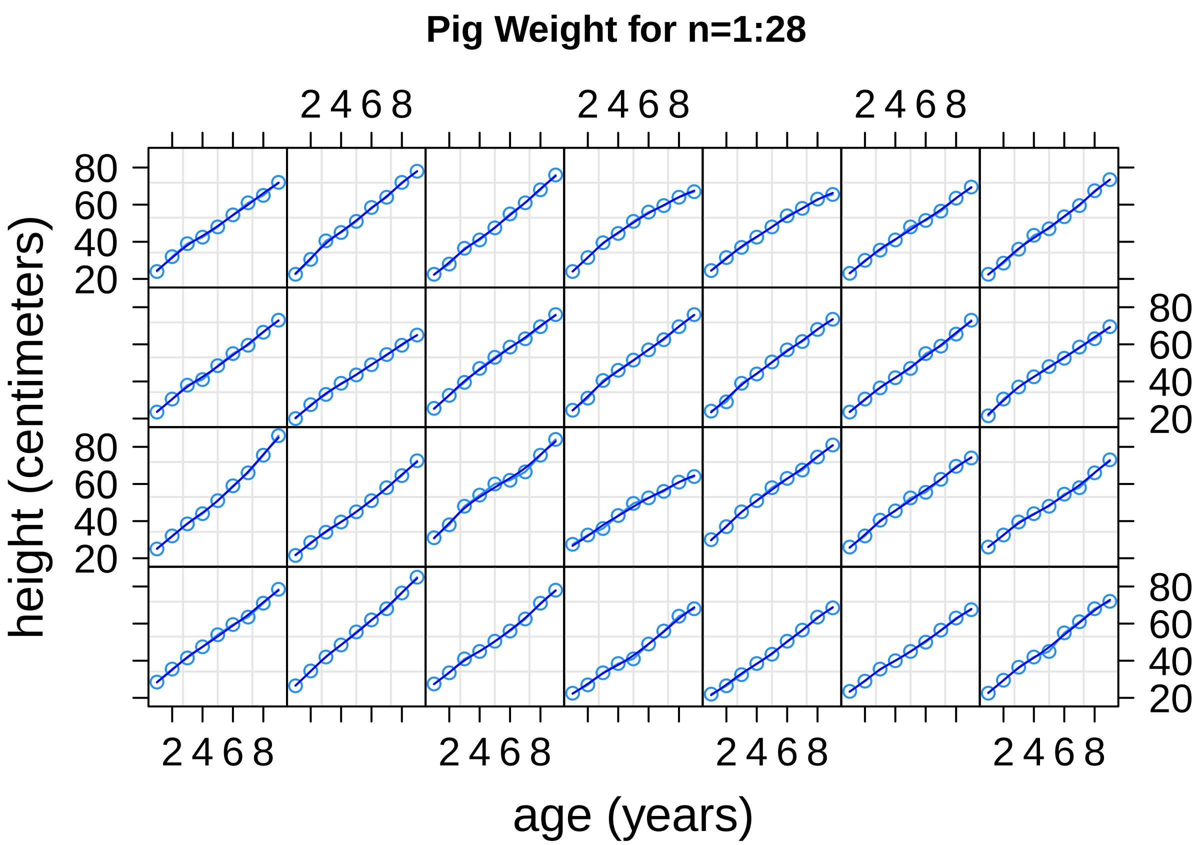
Question 3

Plot the individual BLUP estimates for the first 28 pigs.

```
# reduce the data to just the first 28 pigs
pigs_28 <- pigs %>%
  dplyr::filter(id.num <= 28)

id_num_28 <- pigs_28$id.num
num_weeks_28 <- pigs_28$num.weeks
pig_weight_28 <- pigs_28$weight
```

```
figure <- xyplot(pig_weight_28 ~ num_weeks_28 | id_num_28
                 , groups=id_num_28
                 , strip=F
                 , scales=list(cex=1.25)
                 , xlab=list("age (years)", cex=1.5)
                 , ylab=list("height (centimeters)", cex=1.5)
                 , as.table=T
                 , main="Pig Weight for n=1:28"
#                 , layout=c(4,7)
                 , panel=function(x, y, subscripts, groups)
                     {
                        panel.grid()
                        adolNum  <- id_num_28[subscripts][1]
                        panel.superpose(x, y, subscripts, groups
                                        , col="dodgerblue", type="b"
```

```
                                )
                        panel.xyplot(num_weeks_g, curvEsts[[adolNum]]
                                     , col="blue", type="l"
                                     )
                      }
                  )
figure
```



Question 4

Rerun problem 1 for separate time periods, first weeks 1-5, and then for weeks 6-9.

```
# first split out the data
pigs_epoch_1 <- pigs %>%
  dplyr::filter(num.weeks < 6)
pigs_epoch_2 <- pigs %>%
  dplyr::filter(num.weeks >= 6)
```

First, weeks 1-5.

```
id_num_1 <- pigs_epoch_1$id.num
num_weeks_1 <- pigs_epoch_1$num.weeks
pig_weight_1 <- pigs_epoch_1$weight

fit_epoch_1 <- random_func_mod(id_num=id_num_1, num_weeks=num_weeks_1, pig_weight=pig_weight_1)
```

```
## Warning in lme.formula(pig_weight ~ num_weeks, data = gd, random = Zblock):
```

```
## fewer observations than random effects in all level 3 groups
```

Now, weeks 6-9.

```r
id_num_2 <- pigs_epoch_2$id.num
num_weeks_2 <- pigs_epoch_2$num.weeks
pig_weight_2 <- pigs_epoch_2$weight

fit_epoch_2 <- random_func_mod(id_num=id_num_2, num_weeks=num_weeks_2, pig_weight=pig_weight_2)
```

```
## Warning in lme.formula(pig_weight ~ num_weeks, data = gd, random = Zblock):
## fewer observations than random effects in all level 3 groups
```

Print General Summaries of each fit.

```r
summary(fit_epoch_1)
```

```
## Linear mixed-effects model fit by REML
##  Data: gd
##        AIC      BIC    logLik
##   876.0232 903.8014 -430.0116
##
## Random effects:
##  Formula: ~-1 + Zgbl | dummyId
##  Structure: Multiple of an Identity
##            Zgbl1    Zgbl2    Zgbl3    Zgbl4    Zgbl5    Zgbl6    Zgbl7
## StdDev: 1.167575 1.167575 1.167575 1.167575 1.167575 1.167575 1.167575
##            Zgbl8    Zgbl9   Zgbl10   Zgbl11   Zgbl12   Zgbl13   Zgbl14
## StdDev: 1.167575 1.167575 1.167575 1.167575 1.167575 1.167575 1.167575
##           Zgbl15   Zgbl16   Zgbl17   Zgbl18   Zgbl19   Zgbl20   Zgbl21
## StdDev: 1.167575 1.167575 1.167575 1.167575 1.167575 1.167575 1.167575
##           Zgbl22
## StdDev: 1.167575
##
##  Formula: ~num_weeks | id_num %in% dummyId
##  Structure: General positive-definite
##             StdDev    Corr
## (Intercept) 2.1931526 (Intr)
## num_weeks   0.7343282 0.092
##
##  Formula: ~-1 + Zgrp | id_num %in% id_num %in% dummyId
##  Structure: Multiple of an Identity
##             Zgrp1     Zgrp2     Zgrp3     Zgrp4     Zgrp5     Zgrp6
## StdDev: 0.5823648 0.5823648 0.5823648 0.5823648 0.5823648 0.5823648
##             Zgrp7     Zgrp8     Zgrp9    Zgrp10    Zgrp11    Zgrp12
## StdDev: 0.5823648 0.5823648 0.5823648 0.5823648 0.5823648 0.5823648
##          Residual
## StdDev: 0.6898007
##
## Fixed effects: pig_weight ~ num_weeks
##                 Value Std.Error  DF  t-value p-value
## (Intercept) 19.228154 0.3439811 191 55.89887       0
## num_weeks    6.291623 0.1137356 191 55.31797       0
##  Correlation:
##           (Intr)
## num_weeks -0.052
```

```
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.49191057 -0.42059708  0.03483533  0.46725393  2.03669452
##
## Number of Observations: 240
## Number of Groups:
##                          dummyId               id_num %in% dummyId
##                                1                                48
## id_num.1 %in% id_num %in% dummyId
##                                48
```

```r
summary(fit_epoch_2)
```

```
## Linear mixed-effects model fit by REML
##  Data: gd
##        AIC      BIC    logLik
##   827.4657 853.4418 -405.7328
##
## Random effects:
##  Formula: ~-1 + Zgbl | dummyId
##  Structure: Multiple of an Identity
##            Zgbl1    Zgbl2    Zgbl3    Zgbl4    Zgbl5    Zgbl6    Zgbl7
## StdDev: 1.089982 1.089982 1.089982 1.089982 1.089982 1.089982 1.089982
##            Zgbl8    Zgbl9   Zgbl10   Zgbl11   Zgbl12   Zgbl13   Zgbl14
## StdDev: 1.089982 1.089982 1.089982 1.089982 1.089982 1.089982 1.089982
##           Zgbl15   Zgbl16   Zgbl17   Zgbl18   Zgbl19   Zgbl20   Zgbl21
## StdDev: 1.089982 1.089982 1.089982 1.089982 1.089982 1.089982 1.089982
##           Zgbl22
## StdDev: 1.089982
##
##  Formula: ~num_weeks | id_num %in% dummyId
##  Structure: General positive-definite
##             StdDev    Corr
## (Intercept) 5.9614416 (Intr)
## num_weeks   0.9994226 -0.729
##
##  Formula: ~-1 + Zgrp | id_num %in% id_num %in% dummyId
##  Structure: Multiple of an Identity
##            Zgrp1    Zgrp2    Zgrp3    Zgrp4    Zgrp5    Zgrp6    Zgrp7
## StdDev: 1.053203 1.053203 1.053203 1.053203 1.053203 1.053203 1.053203
##            Zgrp8    Zgrp9   Zgrp10   Zgrp11   Zgrp12  Residual
## StdDev: 1.053203 1.053203 1.053203 1.053203 1.053203 0.7371322
##
## Fixed effects: pig_weight ~ num_weeks
##                  Value Std.Error  DF  t-value p-value
## (Intercept) 17.981193 0.9832818 143 18.28692       0
## num_weeks    6.384294 0.1573653 143 40.56989       0
##  Correlation:
##           (Intr)
## num_weeks -0.777
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -1.762551250 -0.423428607  0.005639395  0.395661740  2.283109789
```

```
##
## Number of Observations: 192
## Number of Groups:
##                            dummyId               id_num %in% dummyId
##                                  1                                48
## id_num.1 %in% id_num %in% dummyId
##                                 48
```

Looking specifically at the between-subject variability for these time periods:

```
print("Weeks 1-5 between-subject variability")
```

```
## [1] "Weeks 1-5 between-subject variability"
```

```
print(fit_epoch_1$modelStruct$reStruct[3])
```

```
## Random effects:
##  Formula: ~-1 + Zgbl | dummyId
##  Structure: Multiple of an Identity
##            Zgbl1    Zgbl2    Zgbl3    Zgbl4    Zgbl5    Zgbl6    Zgbl7
## StdDev: 1.692627 1.692627 1.692627 1.692627 1.692627 1.692627 1.692627
##            Zgbl8    Zgbl9   Zgbl10   Zgbl11   Zgbl12   Zgbl13   Zgbl14
## StdDev: 1.692627 1.692627 1.692627 1.692627 1.692627 1.692627 1.692627
##           Zgbl15   Zgbl16   Zgbl17   Zgbl18   Zgbl19   Zgbl20   Zgbl21
## StdDev: 1.692627 1.692627 1.692627 1.692627 1.692627 1.692627 1.692627
##           Zgbl22 Residual
## StdDev: 1.692627        1
```

```
print("Weeks 6-9 between-subject variability")
```

```
## [1] "Weeks 6-9 between-subject variability"
```

```
print(fit_epoch_2$modelStruct$reStruct[3])
```

```
## Random effects:
##  Formula: ~-1 + Zgbl | dummyId
##  Structure: Multiple of an Identity
##           Zgbl1   Zgbl2   Zgbl3   Zgbl4   Zgbl5   Zgbl6   Zgbl7   Zgbl8
## StdDev: 1.47868 1.47868 1.47868 1.47868 1.47868 1.47868 1.47868 1.47868
##           Zgbl9  Zgbl10  Zgbl11  Zgbl12  Zgbl13  Zgbl14  Zgbl15  Zgbl16
## StdDev: 1.47868 1.47868 1.47868 1.47868 1.47868 1.47868 1.47868 1.47868
##          Zgbl17  Zgbl18  Zgbl19  Zgbl20  Zgbl21  Zgbl22 Residual
## StdDev: 1.47868 1.47868 1.47868 1.47868 1.47868 1.47868        1
```

I find the between-subject variability to be greater for model 1, but not dramatically.