AIC Practice Problems

Your Name Goes Here

Data

These are salamander capture data from Great Smoky Mountains National Park along an elevational gradient. These are counts of *Eurycea wilderae*, the Blue-Ridge Two-Lined Salamander, on transects (25 m long by 2 m wide) and associated environmental and habitat conditions. Transects were in clusters by plot. You can use counts or the presence-absence data (counts coverted to 1 or 0 in the `presence` column).

```
sal <- read.csv("Data/ewilderae.csv", stringsAsFactors = FALSE, header = TRUE)

str(sal)
```

```
## 'data.frame':    254 obs. of  15 variables:
##  $ species          : chr  "EWIL" "EWIL" "EWIL" "EWIL" ...
##  $ Plot_Trans_Survey: chr  "1-1-6/18/2012" "1-1-7/12/2012" "1-2-6/16/2012" "1-2-6/18/2012" ...
##  $ plot             : chr  "1" "1" "1" "1" ...
##  $ transect         : int  1 1 2 2 2 6 1 1 1 2 ...
##  $ Observer         : chr  "JAC" "WEP" "JAC" "JAC" ...
##  $ count            : int  1 1 1 1 2 1 2 1 1 1 ...
##  $ humidity         : num  83.9 87.8 88.3 83.9 87.6 83.9 93.4 91.3 80.4 91.3 ...
##  $ temp             : num  18.5 21.4 20 18.5 20 18.5 16.3 17.6 20.2 17.6 ...
##  $ precip           : num  1.6 17.54 2.42 1.6 3.32 ...
##  $ elev             : num  608 608 609 609 609 ...
##  $ slope            : num  4.29 4.29 6.15 6.15 6.15 ...
##  $ twi              : num  -2.87 -2.87 -7.23 -7.23 -7.23 ...
##  $ lat              : num  35.5 35.5 35.5 35.5 35.5 ...
##  $ lon              : num  -83.3 -83.3 -83.3 -83.3 -83.3 ...
##  $ presence         : int  1 1 1 1 1 1 1 1 1 1 ...
```

```
unique(sal$plot)
```

```
##  [1] "1"   "10"  "11"  "13"  "14"  "15"  "16"  "18"  "19"  "21A" "21B"
## [12] "24"  "25A" "25B" "26"  "27"  "28"  "29"  "3"   "30"  "32"  "33"
## [23] "36"  "38"  "39"  "40"  "41"  "44"  "45"  "46"  "48"  "49"  "5"
## [34] "50"  "51"  "60"  "61"  "7"   "8"   "8B"  "9"
```

```
length(unique(sal$plot)) # number of different plots
```

```
## [1] 41
```

Standardize the Continuous Independent Variables

Turn into a Z-score by substracting the mean from each observeration and dividing by the SD.

Now 95% of your observations will be between -2 and 2 (if your data were normally distributed)

$$x_i' = \frac{x_i - \bar{x}}{\sigma}$$

```
# check the range
range(sal$elev)
```

```
## [1]  446.5692 2020.4159
```

```
range(sal$temp)
```

```
## [1] 12.6 23.0
```

```
# Use the scale function
# ?scale
sal$elev_s <- as.numeric(scale(sal$elev)) # without the as.numeric it creates a form that doesn't play

# by hand
sal$temp_s <- (sal$temp - mean(sal$temp, na.rm = T)) / sd(sal$temp, na.rm = T)

# make a custom function that standardizes the variable.
# Requires a vector and outputs a standardized version of the vector
std <- function(var) {
  var_s <- (var - mean(var, na.rm = T)) / sd(var, na.rm = T)
  return(var_s)
}

# use custom function on a variable
sal$humidity_s <- std(sal$humidity)

# use custom function in combination with dplyr mutate function
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.3.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
sal <- sal %>%
  mutate(precip_s = std(precip),
         twi_s = std(twi),
         slope_s = std(slope))
```

```
## Warning: package 'bindrcpp' was built under R version 3.3.2
```

```
# look at str and summary
str(sal)
```

```
## 'data.frame':    254 obs. of  21 variables:
##  $ species          : chr  "EWIL" "EWIL" "EWIL" "EWIL" ...
##  $ Plot_Trans_Survey: chr  "1-1-6/18/2012" "1-1-7/12/2012" "1-2-6/16/2012" "1-2-6/18/2012" ...
##  $ plot             : chr  "1" "1" "1" "1" ...
##  $ transect         : int  1 1 2 2 2 6 1 1 1 2 ...
##  $ Observer         : chr  "JAC" "WEP" "JAC" "JAC" ...
##  $ count            : int  1 1 1 1 2 1 2 1 1 1 ...
##  $ humidity         : num  83.9 87.8 88.3 83.9 87.6 83.9 93.4 91.3 80.4 91.3 ...
##  $ temp             : num  18.5 21.4 20 18.5 20 18.5 16.3 17.6 20.2 17.6 ...
##  $ precip           : num  1.6 17.54 2.42 1.6 3.32 ...
```

```
##  $ elev             : num  608 608 609 609 609 ...
##  $ slope            : num  4.29 4.29 6.15 6.15 6.15 ...
##  $ twi              : num  -2.87 -2.87 -7.23 -7.23 -7.23 ...
##  $ lat              : num  35.5 35.5 35.5 35.5 35.5 ...
##  $ lon              : num  -83.3 -83.3 -83.3 -83.3 -83.3 ...
##  $ presence         : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ elev_s           : num  -1.31 -1.31 -1.31 -1.31 -1.31 ...
##  $ temp_s           : num  0.041 1.354 0.72 0.041 0.72 ...
##  $ humidity_s       : num  -0.596 0.163 0.261 -0.596 0.124 ...
##  $ precip_s         : num  -0.735 1.328 -0.629 -0.735 -0.512 ...
##  $ twi_s            : num  2.1137 2.1137 -0.0242 -0.0242 -0.0242 ...
##  $ slope_s          : num  -1.37 -1.37 -1.13 -1.13 -1.13 ...
```

```
summary(sal)
```

```
##    species          Plot_Trans_Survey      plot              transect
##  Length:254         Length:254         Length:254         Min.   :1.000
##  Class :character   Class :character   Class :character   1st Qu.:1.000
##  Mode  :character   Mode  :character   Mode  :character   Median :2.000
##                                                           Mean   :2.268
##                                                           3rd Qu.:3.000
##                                                           Max.   :7.000
##
##    Observer             count          humidity          temp
##  Length:254         Min.   :0.000   Min.   :71.40   Min.   :12.60
##  Class :character   1st Qu.:1.000   1st Qu.:83.70   1st Qu.:17.00
##  Mode  :character   Median :1.000   Median :87.60   Median :18.80
##                     Mean   :1.465   Mean   :86.96   Mean   :18.41
##                     3rd Qu.:2.000   3rd Qu.:91.20   3rd Qu.:19.80
##                     Max.   :8.000   Max.   :98.00   Max.   :23.00
##
##      precip            elev            slope             twi
##  Min.   : 0.000   Min.   : 446.6   Min.   : 2.048   Min.   :-9.628
##  1st Qu.: 1.853   1st Qu.: 842.8   1st Qu.: 9.557   1st Qu.:-8.602
##  Median : 4.051   Median :1048.9   Median :13.791   Median :-7.861
##  Mean   : 7.277   Mean   :1109.4   Mean   :15.269   Mean   :-7.182
##  3rd Qu.: 9.429   3rd Qu.:1263.7   3rd Qu.:20.174   3rd Qu.:-6.261
##  Max.   :26.928   Max.   :2020.4   Max.   :52.064   Max.   : 1.093
##  NA's   :1
##       lat             lon            presence          elev_s
##  Min.   :35.50   Min.   :-83.53   Min.   :0.0000   Min.   :-1.7293
##  1st Qu.:35.59   1st Qu.:-83.47   1st Qu.:1.0000   1st Qu.:-0.6955
##  Median :35.59   Median :-83.40   Median :1.0000   Median :-0.1577
##  Mean   :35.60   Mean   :-83.41   Mean   :0.9961   Mean   : 0.0000
##  3rd Qu.:35.61   3rd Qu.:-83.37   3rd Qu.:1.0000   3rd Qu.: 0.4027
##  Max.   :35.69   Max.   :-83.30   Max.   :1.0000   Max.   : 2.3769
##
##      temp_s          humidity_s         precip_s           twi_s
##  Min.   :-2.6308   Min.   :-3.0303   Min.   :-0.9421   Min.   :-1.1992
##  1st Qu.:-0.6383   1st Qu.:-0.6350   1st Qu.:-0.7023   1st Qu.:-0.6960
##  Median : 0.1769   Median : 0.1244   Median :-0.4177   Median :-0.3330
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.6297   3rd Qu.: 0.8255   3rd Qu.: 0.2786   3rd Qu.: 0.4519
##  Max.   : 2.0788   Max.   : 2.1497   Max.   : 2.5440   Max.   : 4.0571
##                                      NA's   :1
```

```
##     slope_s
##  Min.   :-1.6441
##  1st Qu.:-0.7103
##  Median :-0.1837
##  Mean   : 0.0000
##  3rd Qu.: 0.6100
##  Max.   : 4.5760
##
```

Models

Transects within plots were part of the study design so plots should be included as random groups

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
glmer1 <- glmer(count ~ 1 + (1 | plot), data = sal, family = "poisson")
glmer2 <- glmer(count ~ 1 + temp_s + humidity_s + (1 | plot), data = sal, family = "poisson")
glmer3 <- glmer(count ~ 1 + elev_s + (1 | plot), data = sal, family = "poisson")
glmer4 <- glmer(count ~ 1 + elev_s + I(elev_s^2) + (1 | plot), data = sal, family = "poisson")
glmer5 <- glmer(count ~ 1 + slope_s + twi_s + (1 | plot), data = sal, family = "poisson")
glmer6 <- glmer(count ~ 1 + temp_s + humidity_s + elev_s + I(elev_s^2) + slope_s + twi_s + (1 | plot),
```

AICc comparison

```
library(AICcmodavg)
```

```
## Warning: package 'AICcmodavg' was built under R version 3.3.2
```

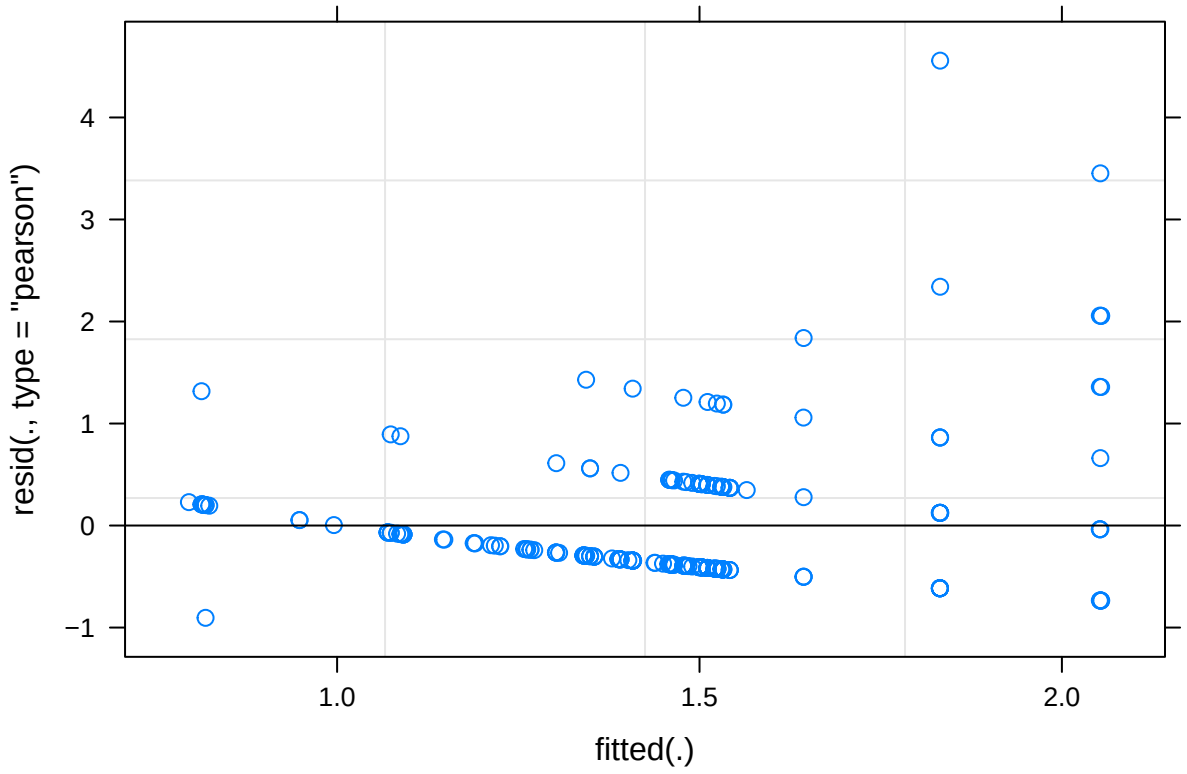
```
##set up candidate models
Cand.mod <- list()

##global model
Cand.mod[[1]] <- glmer1
Cand.mod[[2]] <- glmer2
Cand.mod[[3]] <- glmer3
Cand.mod[[4]] <- glmer4
Cand.mod[[5]] <- glmer5
Cand.mod[[6]] <- glmer6

##assign names to each model
Modnames <- c("intercept only",
              "environ conditions",
              "elevation",
              "quadratic elevation",
              "physical environ",
              "global model")

##model selection table based on AICc
aictab(cand.set = Cand.mod, modnames = Modnames)
```

```
##
## Model selection based on AICc:
```

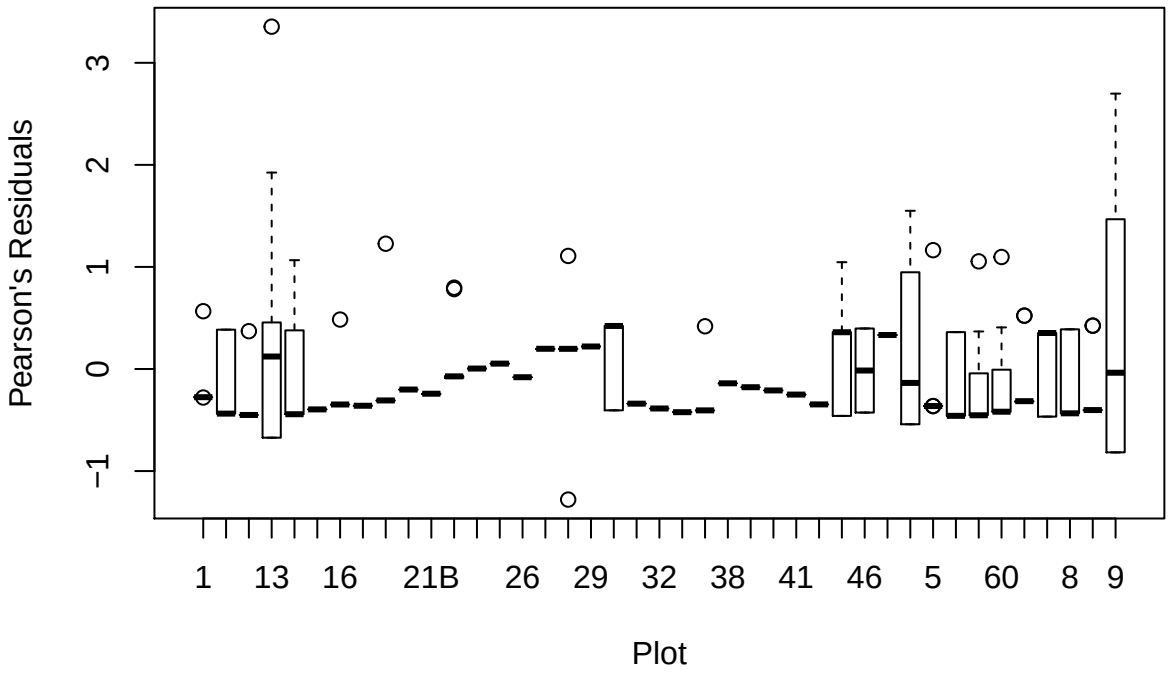
```
##
##                       K   AICc Delta_AICc AICcWt Cum.Wt      LL
## quadratic elevation   4 674.16       0.00   0.54   0.54 -333.00
## elevation             3 676.11       1.95   0.20   0.74 -335.01
## intercept only        2 676.28       2.12   0.19   0.93 -336.11
## environ conditions    4 679.96       5.80   0.03   0.96 -335.90
## physical environ      4 680.27       6.11   0.03   0.98 -336.06
## global model          8 681.14       6.98   0.02   1.00 -332.27
```

Check model fit

```
plot(glmer4)
```




```
boxplot(resid(glmer4) ~ sal$plot,
        xlab = "Plot",
        ylab = "Pearson's Residuals")
```



Random grouping needed?

Because the grouping was part of the experimental design, it is important to always include that in the analysis. However, if it wasn't part of the design but we thought their might be a reason to group certain plots we could still do the same thing as above. In that case, it might be worth seeing if it was worth the grouping.

Compare the best model from above with the same fixed effects model (same model just without any random effects).

```
library(lme4)
glm4 <- glm(count ~ 1 + elev_s + I(elev_s^2), data = sal, family = "poisson")

##set up candidate models
Cand.mod2 <- list()

##global model
Cand.mod2[[1]] <- glmer4
Cand.mod2[[2]] <- glm4

##assign names to each model
Modnames2 <- c("GLMM",
               "GLM")

##model selection table based on AICc
# aictab(cand.set = Cand.mod2, modnames = Modnames2) # throws error

df <- data.frame(model = c("GLMM", "GLM"), AICc = c(AICc(glmer4), AICc(glm4)))
df[order(df$AICc), ]
```

```
##   model     AICc
## 1  GLMM 674.1596
```

2 GLM 674.1915

Grad Student Optional: Using Model Averaging for Multimodel Inference