Forest_Fire_logistic_regression

Bowei Zhang

2023-03-27

```
# delete 'area' column
df <- df %>%
  dplyr::select(-area)
```

model fitting

Full Model Backward Elimination

```
mod.fit.full <- logistf(area_size ~ X + Y + month + day + FFMC + DMC + DC + ISI + temp + RH + wind + ra
                data = df,
                family = binomial)
backward(mod.fit.full, slstay = 0.20)
```

```
## Step  0 : starting model
## Step  1 : removed  day  (P= 0.9766555 )
## Step  2 : removed  RH  (P= 0.9943674 )
## Step  3 : removed  DMC  (P= 0.8610236 )
## Step  4 : removed  ISI  (P= 0.8070721 )
## Step  5 : removed  rain  (P= 0.7190787 )
## Step  6 : removed  DC  (P= 0.5950816 )
## Step  7 : removed  temp  (P= 0.5357923 )
## Step  8 : removed  wind  (P= 0.2436175 )

## logistf(formula = area_size ~ X + Y + month + FFMC, data = df,
##     family = binomial)
## Model fitted by Penalized ML
## Confidence intervals and p-values by Profile Likelihood
##
## Coefficients:
## (Intercept)          X2          X3          X4          X5          X6
## -5.89228418  0.26306068 -1.35416075 -0.49742692 -0.67929097 -0.09144578
##          X7          X8          X9          Y3          Y4          Y5
## -0.55261065  0.78844232  1.18421619  1.19546526  1.52629530  1.30039378
##          Y6          Y8          Y9    monthfeb    monthmar    monthapr
##  0.38323540  1.16046870 -0.52631064  2.07350983  1.17343764  1.40630169
##    monthmay    monthjun    monthjul    monthaug    monthsep    monthoct
##  1.69432113  1.18232922  1.92112039  1.63594208  1.82009651  0.91391050
##    monthnov    monthdec        FFMC
```

```
##  0.73675768  5.24063539  0.03714333
##
## Likelihood ratio test=58.60721 on 26 df, p=0.0002573382, n=517
```

```
mod.fit <- logistf(area_size ~ X + Y + month + FFMC,
               data = df,
               family = binomial)
summary(mod.fit)
```

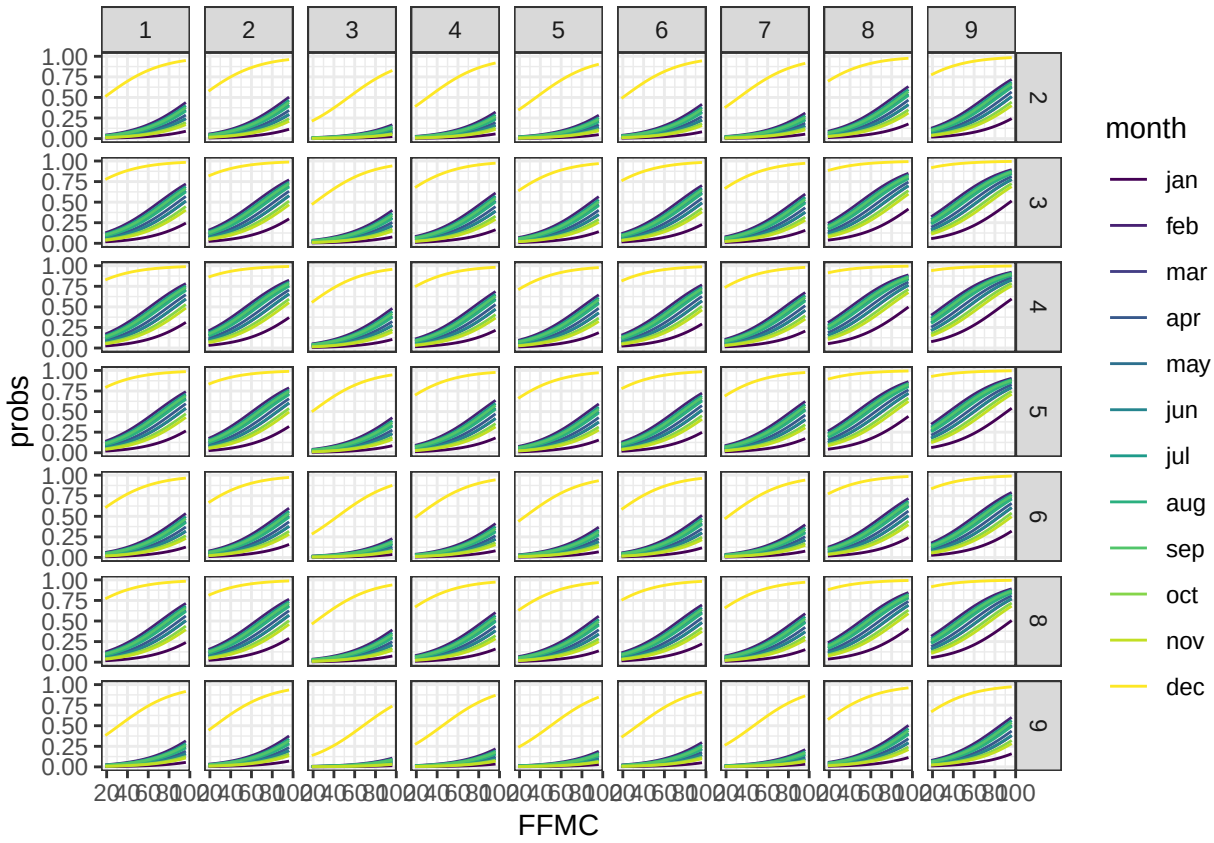
```
## logistf(formula = area_size ~ X + Y + month + FFMC, data = df,
##     family = binomial)
##
## Model fitted by Penalized ML
## Coefficients:
##                     coef   se(coef)    lower 0.95  upper 0.95       Chisq
## (Intercept) -5.89228418 2.54433948 -12.106807340 -1.15747276  6.77039610
## X2           0.26306068 0.39899157  -0.524826520  1.05211677  0.43096284
## X3          -1.35416075 0.46451485  -2.290896303 -0.45520226  8.82862706
## X4          -0.49742692 0.40783836  -1.314114870  0.29638725  1.49956402
## X5          -0.67929097 0.51646297  -1.709572947  0.33073408  1.73435966
## X6          -0.09144578 0.42536310  -0.939163122  0.73916346  0.04606839
## X7          -0.55261065 0.44190928  -1.434979345  0.30793172  1.57625088
## X8           0.78844232 0.59409350  -0.369302795  2.00662400  1.76514094
## X9           1.18421619 1.00910830  -0.645281826  3.59726762  1.52645940
## Y3           1.19546526 0.47347105   0.275998857  2.14152242  6.52465531
## Y4           1.52629530 0.41649591   0.721215111  2.36564399 14.10933574
## Y5           1.30039378 0.43368019   0.461607925  2.17032144  9.34382644
## Y6           0.38323540 0.60022212  -0.839426147  1.56069912  0.39194796
## Y8           1.16046870 1.74951790  -2.015278016  6.24281668  0.49194119
## Y9          -0.52631064 1.28062624  -3.337104103  1.90680548  0.16971224
## monthfeb     2.07350983 1.69996340  -1.214641586  7.14740360  1.58989036
## monthmar     1.17343764 1.67672841  -2.072429581  6.22111709  0.52192790
## monthapr     1.40630169 1.76828480  -2.009971120  6.53535365  0.66653328
## monthmay     1.69432113 2.02008304  -2.238213358  7.06691104  0.72575514
## monthjun     1.18232922 1.73630563  -2.176911601  6.28184349  0.48844124
## monthjul     1.92112039 1.69392588  -1.353226283  6.98182027  1.39907563
## monthaug     1.63594208 1.65990398  -1.577214820  6.66620899  1.06327922
## monthsep     1.82009651 1.65933437  -1.393609845  6.85149994  1.31400019
## monthoct     0.91391050 1.73844219  -2.445569655  6.01239715  0.28753276
## monthnov     0.73675768 2.33594690  -4.901783347  6.35793231  0.09533190
## monthdec     5.24063539 2.20009133   1.288518375 11.20310992  6.63123851
## FFMC         0.03714333 0.02295742  -0.008756432  0.08857138  2.47654181
##                        p method
## (Intercept) 0.0092682262      2
## X2          0.5115168775      2
## X3          0.0029654140      2
## X4          0.2207384571      2
## X5          0.1878550855      2
## X6          0.8300515199      2
## X7          0.2093015512      2
## X8          0.1839845715      2
## X9          0.2166446051      2
## Y3          0.0106389165      2
## Y4          0.0001724856      2
```

```
## Y5          0.0022373788      2
## Y6          0.5312772459      2
## Y8          0.4830626608      2
## Y9          0.6803676726      2
## monthfeb    0.2073417990      2
## monthmar    0.4700204498      2
## monthapr    0.4142628795      2
## monthmay    0.3942625818      2
## monthjun    0.4846234585      2
## monthjul    0.2368784017      2
## monthaug    0.3024684614      2
## monthsep    0.2516716747      2
## monthoct    0.5918059508      2
## monthnov    0.7575051367      2
## monthdec    0.0100205568      2
## FFMC        0.1155560782      2
##
## Method: 1-Wald, 2-Profile penalized log-likelihood, 3-None
##
## Likelihood ratio test=58.60721 on 26 df, p=0.0002573382, n=517
## Wald test = 46.74999 on 26 df, p = 0.007510991
```

Visualization

```
# Create a new dataset with predicted values
pred_probs <- with(
  df,
  expand.grid(
    X = unique(X),
    Y = unique(Y),
    month = unique(month),
    FFMC = seq(min(FFMC), max(FFMC), length.out = 100)
  )
)
pred_probs$probs <- predict(mod.fit, newdata = pred_probs, type = "response")

# ggplot2
fig1 <- ggplot(pred_probs, aes(x = FFMC, y = probs, col = month)) +
  geom_line() +
  facet_grid(Y ~ X) +
  theme_bw()+
  scale_color_viridis_d()
fig1
```




```
ggsave("../results/fitted_logistic_model.png", fig1, width = 20, height = 12, dpi = 320)
```

```
# Hosmer Lemeshow test
generalhoslem::logitgof(df$area_size, mod.fit$predict)
```

```
##
##  Hosmer and Lemeshow test (binary model)
##
## data:  df$area_size, mod.fit$predict
## X-squared = 5.8354, df = 8, p-value = 0.6657
```

The Hosmer-Lemeshow test has a p-value of 0.67 which fails to reject the null; hence, the model is a good fit.