Hamida Paiman, Vaiswi Patel, Amanda Tsui, Mehaer Chhabda, Emily Gill

Emily: Wildfire.Count

Adding polynomial terms and interactions

```
NOAAGISSWD$Year_sq <- NOAAGISSWD$Year^2
NOAAGISSWD$delta.temp_sq <- NOAAGISSWD$delta.temp^2
NOAAGISSWD$Year_delta.temp <- NOAAGISSWD$Year * NOAAGISSWD$delta.temp
```

Model 1: Basic Model with Only Year

```
model1 <- glm(Wildfire.Count ~ Year, family = binomial(link = "logit"), data = NOAAGISSWD)
summary(model1)
```

```
##
## Call:
## glm(formula = Wildfire.Count ~ Year, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -243.44788   72.16830  -3.373 0.000743 ***
## Year           0.12163    0.03606   3.373 0.000743 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 60.997  on 43  degrees of freedom
## Residual deviance: 43.787  on 42  degrees of freedom
## AIC: 47.787
##
## Number of Fisher Scoring iterations: 4
```

Model 2: Basic Model with Only delta.temp

```
model2 <- glm(Wildfire.Count ~ delta.temp, family = binomial(link = "logit"), data = NOAAGISSWD)
summary(model2)
```

```
##
## Call:
## glm(formula = Wildfire.Count ~ delta.temp, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -3.118      1.033  -3.020  0.00253 **
## delta.temp     5.699      1.799   3.168  0.00154 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 60.997  on 43  degrees of freedom
## Residual deviance: 45.406  on 42  degrees of freedom
## AIC: 49.406
##
## Number of Fisher Scoring iterations: 4
```

Model 3: Model with Both Year and delta.temp

```r
model3 <- glm(Wildfire.Count ~ delta.temp + Year, family = binomial(link = "logit"), data = NOAAGISSWD)
summary(model3)
```

```
##
## Call:
## glm(formula = Wildfire.Count ~ delta.temp + Year, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -212.80370  169.82370  -1.253    0.210
## delta.temp     0.81269    4.12303   0.197    0.844
## Year           0.10610    0.08587   1.236    0.217
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 60.997  on 43  degrees of freedom
## Residual deviance: 43.748  on 41  degrees of freedom
## AIC: 49.748
##
## Number of Fisher Scoring iterations: 4
```

Model 4: Adding Interaction Term

```r
model4 <- glm(Wildfire.Count ~ delta.temp + Year + Year_delta.temp, family = binomial(link = "logit"), data = NOAAGISSWD)
summary(model4)
```

```
##
## Call:
## glm(formula = Wildfire.Count ~ delta.temp + Year + Year_delta.temp,
##     family = binomial(link = "logit"), data = NOAAGISSWD)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -188.74474  218.72821  -0.863    0.388
## delta.temp       -52.78261  306.29915  -0.172    0.863
## Year               0.09409    0.11004   0.855    0.393
## Year_delta.temp    0.02670    0.15263   0.175    0.861
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 60.997  on 43  degrees of freedom
## Residual deviance: 43.717  on 40  degrees of freedom
## AIC: 51.717
##
## Number of Fisher Scoring iterations: 4
```

Model 5: Adding Quadratic Terms (Full Second Order)

```
model5 <- glm(Wildfire.Count ~ delta.temp + Year + Year_delta.temp + Year_sq + delta.temp_sq,
              family = binomial(link = "logit"), data = NOAAGISSWD)
summary(model5)
```

```
##
## Call:
## glm(formula = Wildfire.Count ~ delta.temp + Year + Year_delta.temp +
##     Year_sq + delta.temp_sq, family = binomial(link = "logit"),
##     data = NOAAGISSWD)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      7.634e+04  8.117e+04   0.941    0.347
## delta.temp       4.724e+03  4.300e+03   1.098    0.272
## Year            -7.769e+01  8.227e+01  -0.944    0.345
## Year_delta.temp -2.397e+00  2.179e+00  -1.100    0.271
## Year_sq          1.976e-02  2.084e-02   0.948    0.343
## delta.temp_sq    6.864e+01  5.750e+01   1.194    0.233
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 60.997  on 43  degrees of freedom
## Residual deviance: 41.831  on 38  degrees of freedom
## AIC: 53.831
##
## Number of Fisher Scoring iterations: 6
```

Model Comparison using AIC

```
AIC_values <- c(AIC(model1), AIC(model2), AIC(model3), AIC(model4), AIC(model5))
names(AIC_values) <- c("Model 1 (Year only)", "Model 2 (delta.temp only)",
                       "Model 3 (Year + delta.temp)", "Model 4 (Interaction)",
                       "Model 5 (Full Second-Order)")
print(AIC_values)
```

```
##         Model 1 (Year only)   Model 2 (delta.temp only)
##                    47.78727                    49.40551
## Model 3 (Year + delta.temp)       Model 4 (Interaction)
##                    49.74829                    51.71748
## Model 5 (Full Second-Order)
##                    53.83097
```

PRESS Calculation for Each Model

```r
# Function to calculate PRESS
PRESS_logit <- function(model) {
  residuals <- residuals(model, type = "deviance")
  hat_values <- lm.influence(model)$hat
  press <- sum((residuals / (1 - hat_values))^2)
  return(press)
}

# Calculate PRESS for each model
press_values <- c(PRESS_logit(model1), PRESS_logit(model2),
                  PRESS_logit(model3), PRESS_logit(model4),
                  PRESS_logit(model5))
names(press_values) <- c("Model 1 (Year only)", "Model 2 (delta.temp only)",
                         "Model 3 (Year + delta.temp)", "Model 4 (Interaction)",
                         "Model 5 (Full Second-Order)")
print(press_values)
```

```
##         Model 1 (Year only)   Model 2 (delta.temp only)
##                    47.91866                    49.60191
## Model 3 (Year + delta.temp)       Model 4 (Interaction)
##                    50.51476                    51.91351
## Model 5 (Full Second-Order)
##                    57.48090
```

Bootstrapping for Each Model

```
# Running bootstrapping for each model to estimate the stability of coefficients.
boot_logit_model1 <- function(data, indices) {
  fit <- glm(Wildfire.Count ~ Year, family = binomial, data = data[indices, ])
  return(coef(fit))
}

boot_logit_model2 <- function(data, indices) {
  fit <- glm(Wildfire.Count ~ delta.temp, family = binomial, data = data[indices, ])
  return(coef(fit))
}

boot_logit_model3 <- function(data, indices) {
  fit <- glm(Wildfire.Count ~ delta.temp + Year, family = binomial, data = data[indices, ])
  return(coef(fit))
}

boot_logit_model4 <- function(data, indices) {
  fit <- glm(Wildfire.Count ~ delta.temp + Year + Year_delta.temp, family = binomial, data = data[indices, ])
  return(coef(fit))
}

boot_logit_model5 <- function(data, indices) {
  fit <- glm(Wildfire.Count ~ delta.temp + Year + Year_delta.temp + Year_sq + delta.temp_sq, family = binomial, d
ata = data[indices, ])
  return(coef(fit))
}

# Perform bootstrapping with 1000 replications
set.seed(123)
wildfire_ci_logit_model1 <- boot(data = NOAAGISSWD, statistic = boot_logit_model1, R = 1000)
wildfire_ci_logit_model2 <- boot(data = NOAAGISSWD, statistic = boot_logit_model2, R = 1000)
wildfire_ci_logit_model3 <- boot(data = NOAAGISSWD, statistic = boot_logit_model3, R = 1000)
wildfire_ci_logit_model4 <- boot(data = NOAAGISSWD, statistic = boot_logit_model4, R = 1000)
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
wildfire_ci_logit_model5 <- boot(data = NOAAGISSWD, statistic = boot_logit_model5, R = 1000)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
# Print bootstrapping results
print(wildfire_ci_logit_model1)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = NOAAGISSWD, statistic = boot_logit_model1, R = 1000)
##
##
## Bootstrap Statistics :
##          original       bias    std. error
## t1* -243.4478775 -21.64367711 81.75694052
## t2*    0.1216327   0.01081798  0.04085639
```

```
print(wildfire_ci_logit_model2)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = NOAAGISSWD, statistic = boot_logit_model2, R = 1000)
##
##
## Bootstrap Statistics :
##      original     bias    std. error
## t1* -3.117966 -0.3099106    1.265640
## t2*  5.698753  0.5425598    2.236757
```

```
print(wildfire_ci_logit_model3)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = NOAAGISSWD, statistic = boot_logit_model3, R = 1000)
##
##
## Bootstrap Statistics :
##         original       bias    std. error
## t1* -212.8036980 -31.16104522 254.4431114
## t2*    0.8126896   0.20957533   5.7156988
## t3*    0.1061008   0.01550627   0.1285781
```

```
print(wildfire_ci_logit_model4)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = NOAAGISSWD, statistic = boot_logit_model4, R = 1000)
##
##
## Bootstrap Statistics :
##          original        bias     std. error
## t1* -188.74474055 -149.49140584  3927.930885
## t2*  -52.78260628 -555.16651522 17352.565807
## t3*    0.09409445    0.07561554     1.994381
## t4*    0.02670218    0.27560843     8.609956
```

```
print(wildfire_ci_logit_model5)
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = NOAAGISSWD, statistic = boot_logit_model5, R = 1000)
##
##
## Bootstrap Statistics :
##          original        bias     std. error
## t1*  7.634483e+04  6.603196e+04 7.163235e+06
## t2*  4.723529e+03  1.081315e+04 3.353721e+05
## t3* -7.768639e+01 -6.926153e+01 7.230823e+03
## t4* -2.396904e+00 -5.522072e+00 1.698135e+02
## t5*  1.976133e-02  1.813742e-02 1.824844e+00
## t6*  6.863539e+01  2.203118e+02 4.754929e+03
```

Analysis for Wildfire Section:

Model 1 (Only Year):

Coefficients for both Intercept and Year are highly significant, with p-values < 0.001 (0.000743). This indicates strong evidence that Year is associated with Wildfire.Count.

Model 2 (Only delta.temp):

Both the Intercept and delta.temp coefficients are significant, with p-values < 0.001 (0.00253). This suggests that delta.temp alone is also a plausible predictor.

Model 3 (Year + delta.temp):

Both Year and delta.temp are included, but neither coefficient reaches statistical significance at the 95% level (p > 0.2). This model may not be plausible.

Model 4 (Adding Interaction):

The coefficients for Year, delta.temp, and their interaction term all have p-values > 0.2, indicating weak evidence of association. This model is less plausible.

Model 5 (Full Second-Order):

This model has the highest complexity, including all polynomial and interaction terms. However, none of the coefficients are significant at the 95% level, making it less plausible.

Model Selection using AIC and PRESS

AIC Values:

- Model 1: 47.79 (lowest AIC, suggesting best fit by AIC)
- Model 2: 49.41

- Model 3: 49.75

Higher-order models (Models 4-5) have AICs over 51, indicating worse fit.

<u>PRESS Values:</u>

Similar to AIC, Model 1 has the lowest PRESS score, indicating it best predicts out-of-sample data compared to other models (47.92).

Conclusion:

- Model 1 (Only Year) is the best model based on both statistical significance and model fit (lowest AIC and PRESS). This suggests that Year alone provides a plausible and optimal fit for predicting Wildfire.Count.
- The bootstrapping results provide additional information on coefficient variability. There are convergence issues and the occurrence of fitted probabilities close to 0 or 1. Although there are warnings, we still treat Wildfire.Count as binary in this data set, as it only takes on the values of 0 or 1. This makes the data effectively binary in this instance. For binary outcomes, addressing logistic model convergence issues is generally preferable to switching to linear regression.