

report

Jingning Yang

2/21/2020

Introduction

Our client, Reshma Satish, a PhD student at Boston University's Department of Forensic Anthropology, administered a project involves fracturing 135 long deer bones using a standardized fracturing apparatus over a period of 540 days. The bones are broken into 9 groups with 15 bones in each group and each group is fractured at different times(0 days,30 days,60 days,90 days,180 days,270 days, etc.) Her goal was to investigate how time of exposure of long bones affects variables such as the type of fracture angle, the type of fracture surface morphology, etc. that is created on the bones. Currently, our client only has data until day 270 since this experiment still keep going.

This report will start with a description of the EDA, as well as Data analysis/Data structure. Then we will describe the methods we will use on modeling part, followed by our analysis.

EDA (YW)

Data Cleaning

```
## [1] "integer"
```

Modeling & Interpretation (YW-poisson) (HG-fracture type & surface)

We used logistic regression, poisson regression, and multinomial logit model based on different type of predictors in our dataset.

Bisect Fracture part

Since we assume response variable "bisect fracture" is a binary outcomes(0 or 1) and predictor "days" is continuous, we are using logistic regression to figure out relationship between the probability of "bisect fracture" and value of "days".

In addition, when "bisect fracture" is 0, it means the bone did not break by the standard fracturing machine, otherwise, it means the bone broke by the machine.

```
## (Intercept)    bone$days
## 0.029963710 0.003387971

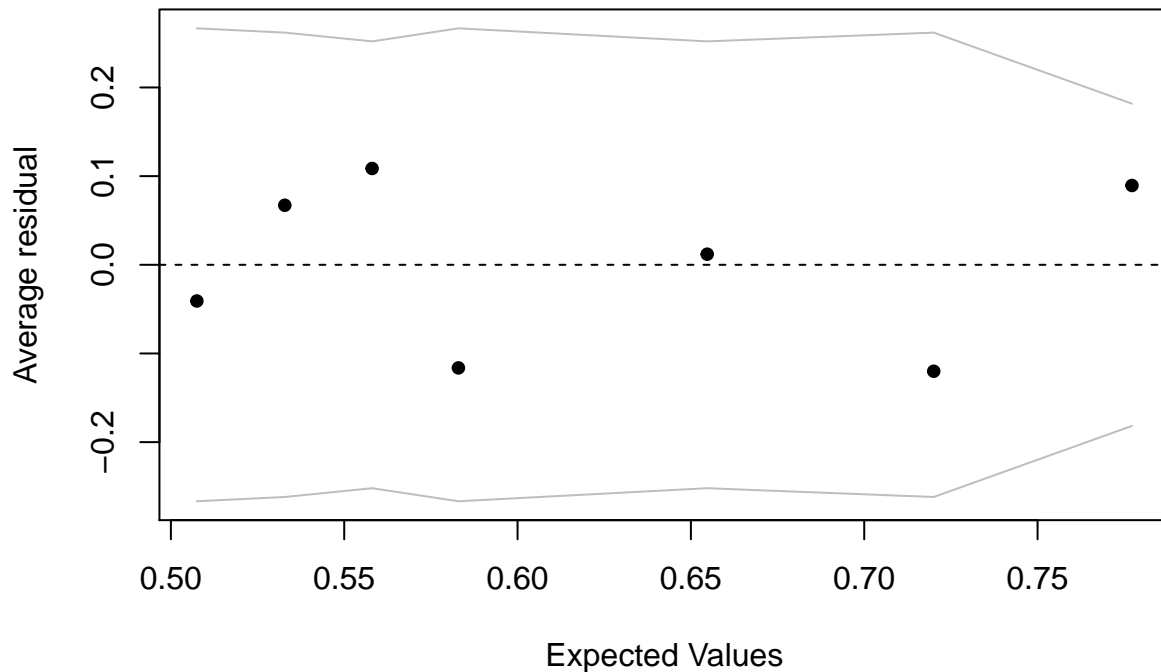
## (Intercept)    bone$days
##    0.9207737    0.0517851
```

From the output, the estimated intercept coefficient 0.03 correspond to the log odds of broke the bone with average days, in other words, on average days, expected probability of broke the bone is $\text{expit}(0.03)=0.4925=49\%$.

And one-unit increase in the predictor "days" increases the log odds of broke the bone (versus not broke the bone) by 0.0034.

Further, p-value of days is 0.052 indicates that it is somewhat not significant in determining whether the bone broke by the machine or not.

Binned residual plot



Since the discrete residual plot does not tell us much about the model fit, therefore, we are using Binned residual plot for this model, due to every points fall within the bounds, our model fit well in this case.

Fracture like part

We assume “Fracture-like” is a binary outcomes(0 or 1) and predictor is “days” as well, therefore, as idea with above part, we used logistic regression to figure out their relationship.

In addition, when “Fracture-like” is 0, it means the bone has no surface feature after cut by the standard fracturing machine, otherwise, it means fracturing machine made a mark on the bone.

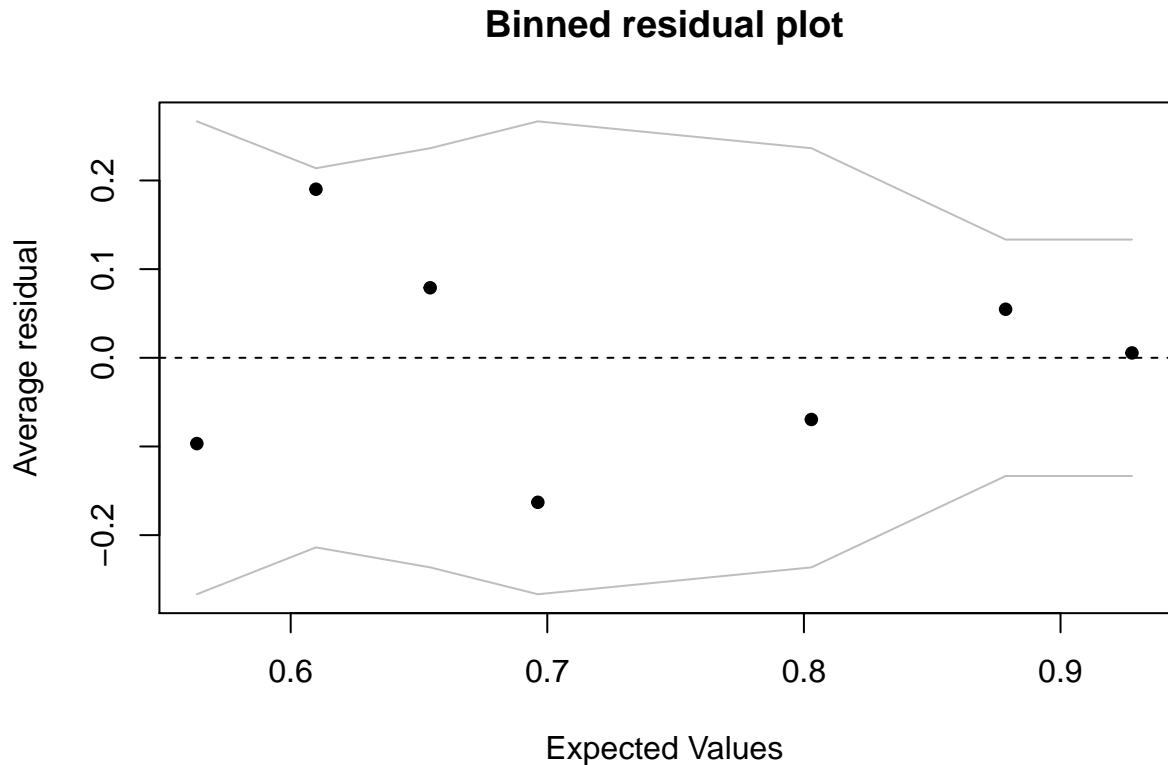
```
## (Intercept)    bone$days
##  0.25509986  0.00638606

## (Intercept)    bone$days
##  0.425653911  0.005468143
```

From the output, the estimated intercept coefficient 0.2551 correspond to the log odds of made a mark on bone with average days, in other words, on average days, expected probability of made a mark on bone is $\text{expit}(0.2551)=0.436=43.6\%$.

And one-unit increase in the predictor “days” increases the log odds of made a mark on bone(versus no surface feature on bone) by 0.0064.

Also, p-value of days is 0.0055 indicates that it is significant in determining whether the bone has any surface feature after cut by the machine.



logic with above part, our model fit well in this case.

Fracture Angle part

We assume “Fracture Angle” as no natural-order variable(nominal), since we cannot compare the order of angle of fresh bone and dry bone, and also N/A(non applicable) in this variable, so we used multinomial logit model in this case.

```
##                (Intercept)    bone$days
## non applicable    2.239716 -0.005672016
## OBLIQUE          2.228370 -0.002974056
## RIGHT ANGLES     -1.214702 -0.007741723
```

From the output, the baseline is Fracture Angle = Intermediate.

A one-unit increase in the variable “days” is associated with the decrease in the log odds of being in Intermediate fracture angle vs. non applicable fracture angle in the amount of 0.0057.

A one-unit increase in the variable “days” is associated with the decrease in the log odds of being in Intermediate fracture angle vs. oblique fracture angle in the amount of 0.003.

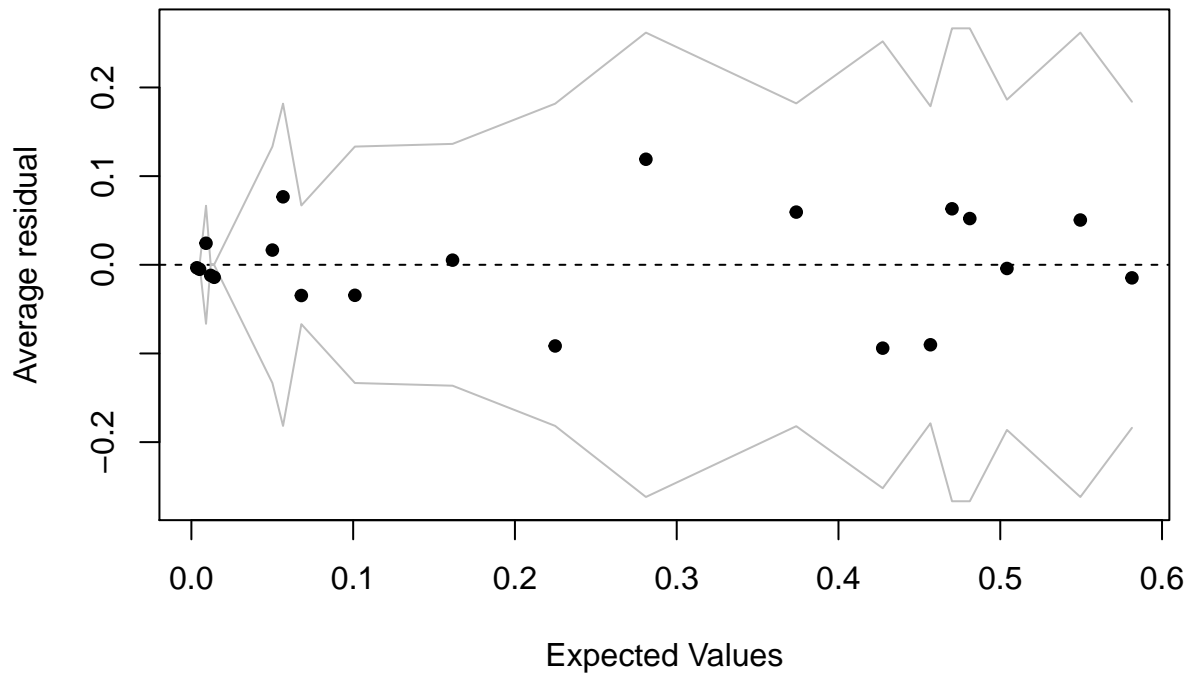
In other words, if one-unit increase in “days”, your chances of staying in the intermediate fracture angle category are higher compared to staying in non applicable and oblique.

Since summary of multinomial logit model does not provide p-values, we can get significance of the coefficients by calculating z score and input it into the formula:

```
##                (Intercept)    bone$days
## non applicable 0.0003749240 0.04724377
## OBLIQUE       0.0003380515 0.26601701
## RIGHT ANGLES  0.4104893086 0.45815716
```

From output, it represents p-value of coefficients. Only the coefficient of non applicable is somewhat significant because of $0.047 < 5\%$.

Binned residual plot



Looks like our model fit kind of well in this case, except there might not good enough when expected values really small.

Number of Fragments part(YW)

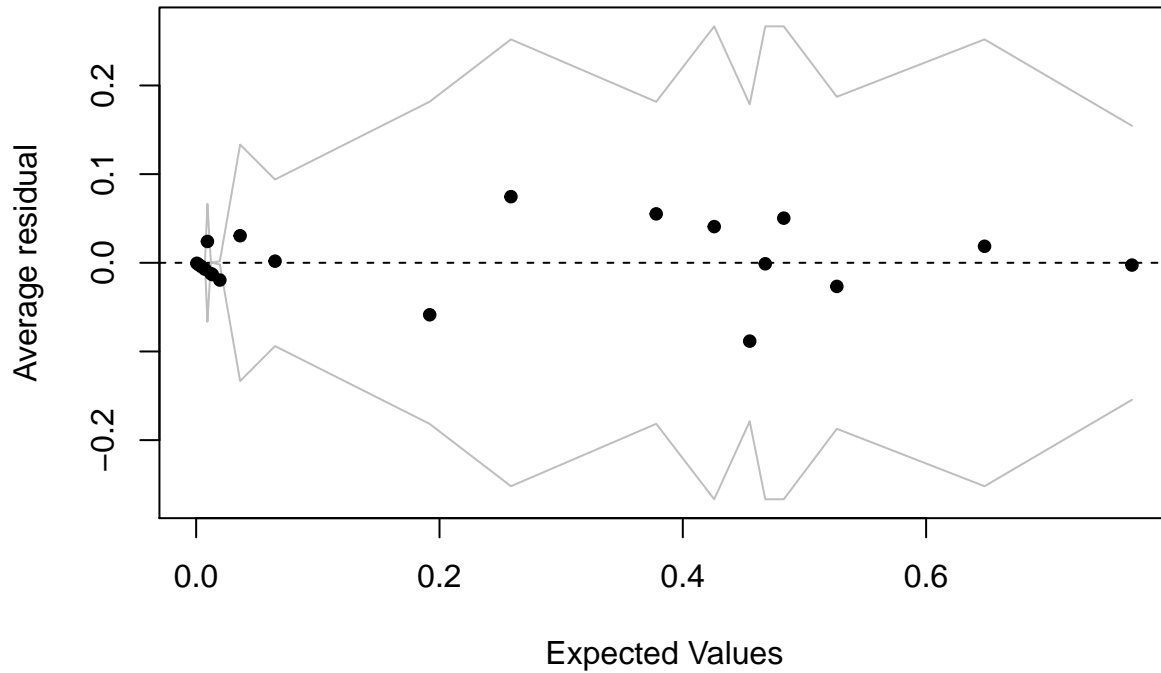
Fracture Outline

We assume “Fracture Outline” as no natural-order(nominal) variable, same reason as “Fracture Angle” part.

```
##          (Intercept)    bone$days
## INTERMEDIATE    -1.7143707 -0.015467533
## non applicable    0.1259323 -0.004329254
## TRANSVERSE     -3.3809875 -0.005924679

##          (Intercept)    bone$days
## INTERMEDIATE    0.02748415 0.18352853
## non applicable    0.68625024 0.01718665
## TRANSVERSE      0.01405271 0.56466321
```

Binned residual plot

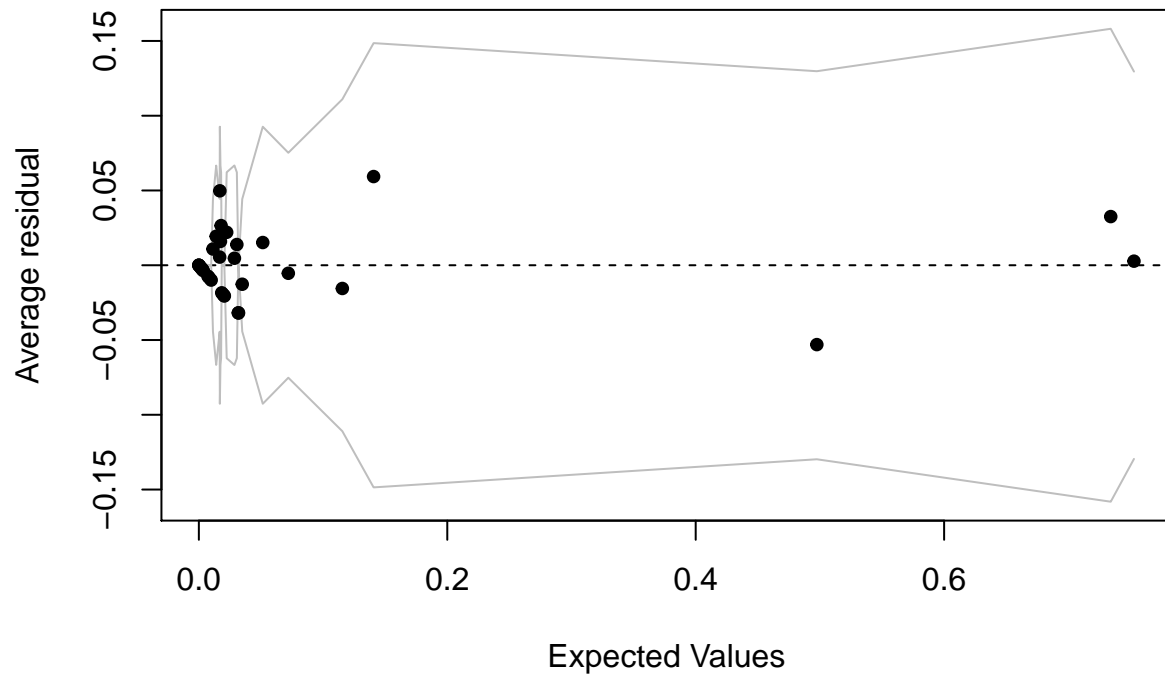


Fractographic Features

```
## Call:
## multinom(formula = bone$`Fractographic Features` ~ bone$days,
##   trace = FALSE)
##
## Coefficients:
##               (Intercept)      bone$days
## BONE HACKLE              -4.513009e+01  1.274155e-01
## BONE HACKLE, WAKE HACKLE, ARREST RIDGE  1.702261e+00 -4.177227e-03
## BONE MIRROR                2.366450e+00 -7.872610e-03
## COMPRESSION CURL           1.830044e+00 -2.915312e-03
## COMPRESSIONAL CURL         2.973164e+00 -2.718052e-02
## No Feature                 6.140032e+00 -8.824577e-03
## WAKE HACKLE                4.578751e+00 -1.119986e-02
## WAKE HACKLE, ARREST RIDGE  -1.268053e-04  4.482885e-07
## WAKE HACKLE, BONE HACKLE    3.054647e+00 -9.973191e-03
## WAKE HACKLE, BONE MIRROR    1.863558e-04 -8.068022e-07
## WAKE HACKLE, WALLNER LINES  2.366409e+00 -7.872100e-03
##
## Std. Errors:
##               (Intercept)      bone$days
## BONE HACKLE              0.0000702144  0.004553559
## BONE HACKLE, WAKE HACKLE, ARREST RIDGE  2.9983264394  0.010720047
## BONE MIRROR                2.8939021781  0.010812737
## COMPRESSION CURL           2.9342264730  0.010284139
## COMPRESSIONAL CURL         2.9629897222  0.027391160
## No Feature                 2.6788598637  0.009242398
## WAKE HACKLE                2.7049689701  0.009596848
```

```
## WAKE HACKLE, ARREST RIDGE      3.7743928967 0.012959267
## WAKE HACKLE, BONE HACKLE      2.8020223189 0.010498466
## WAKE HACKLE, BONE MIRROR      3.7742136698 0.012958896
## WAKE HACKLE, WALLNER LINES    2.8939037642 0.010812636
##
## Residual Deviance: 226.1467
## AIC: 270.1467
```

Binned residual plot



Conclusion