Lab 9

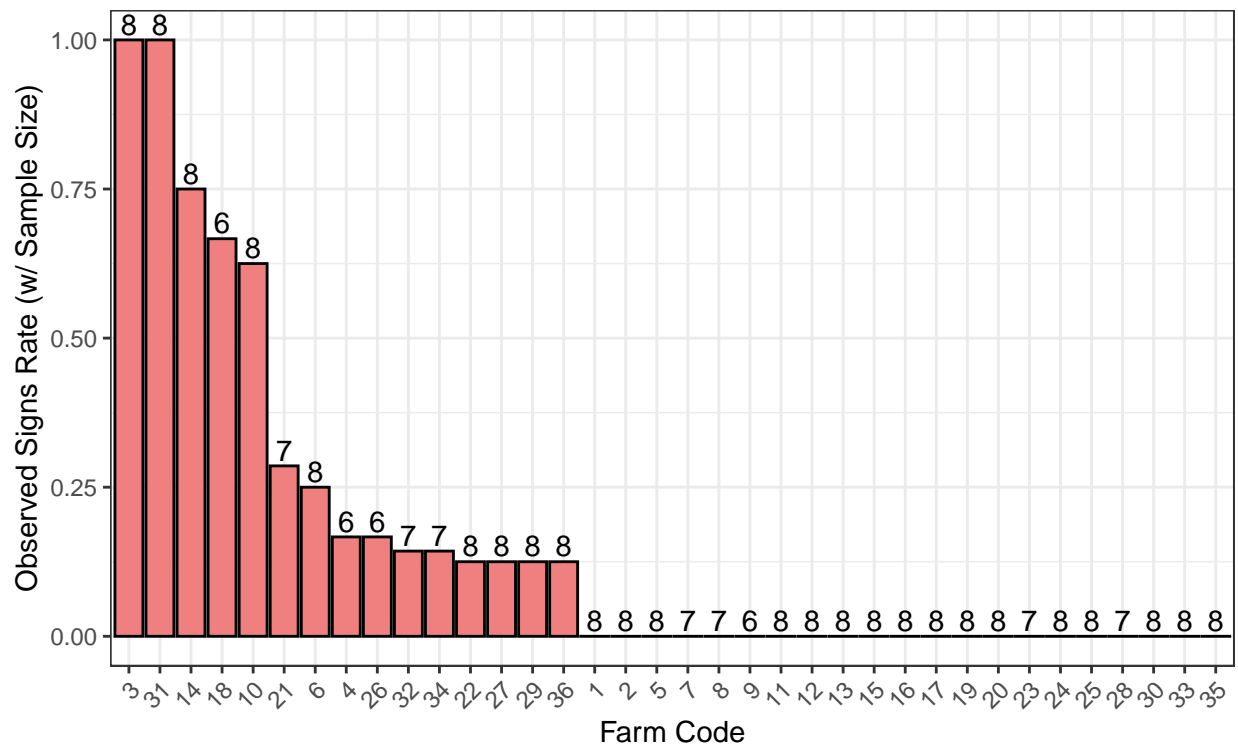Emily Gentles   Weiyi Liu   Jack McCarthy   Qinzhe Wang

10/28/2021

EDA

First a look at the distribution of the response. We can look at the mean **signs_in_yard** value for each farm to get a "rate" of signs being observed.


It seems like the signs rates vary widely across farms, so we should expect to see heterogeneity in our final model. Now we can look at the signs rates for some of the candidate predictors. First, a look at the rates for each value of the number of active badger settlements.
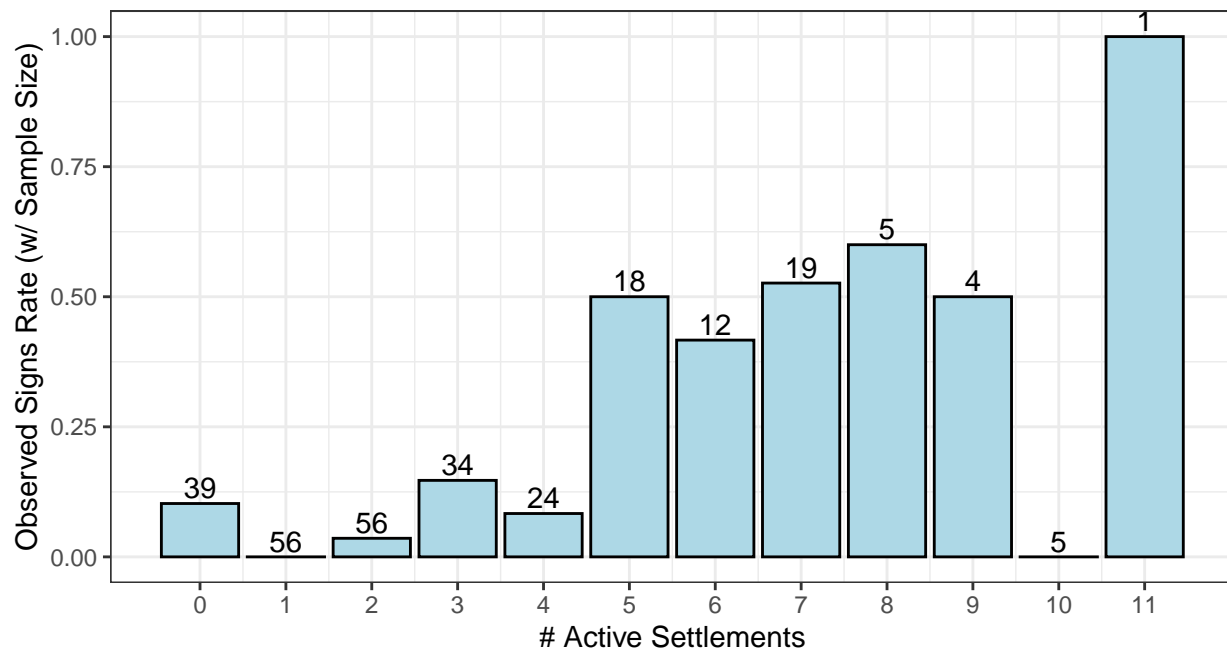
There seems to be a positive relationship between the number of active settlements and the rate of badger signs in the yard. We may also look at the relationship between the number of cattle and the signs rate. To do so we will bin the number of cattle into 10% quantile increments and chart the rate for each bin.
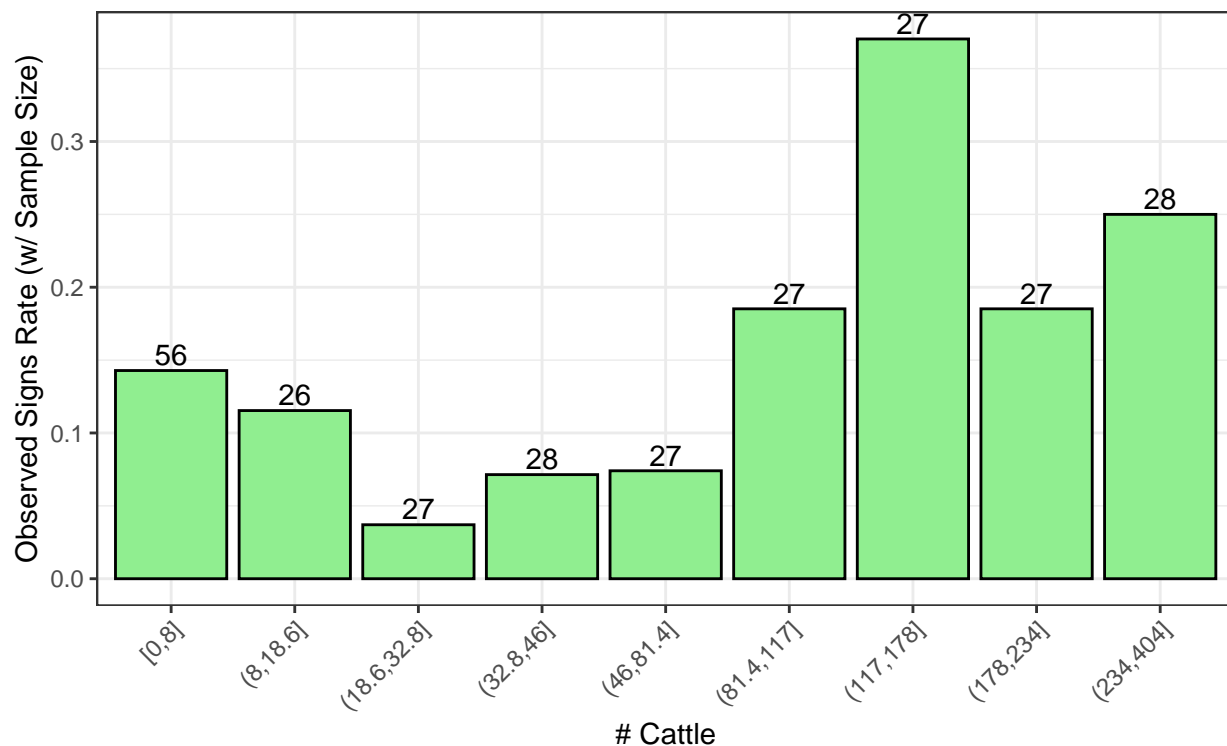
Again, there seems to be a slight positive relationship between the number of cattle and the rate of badger sign observation. The plot below is an alternative view of this relationship.
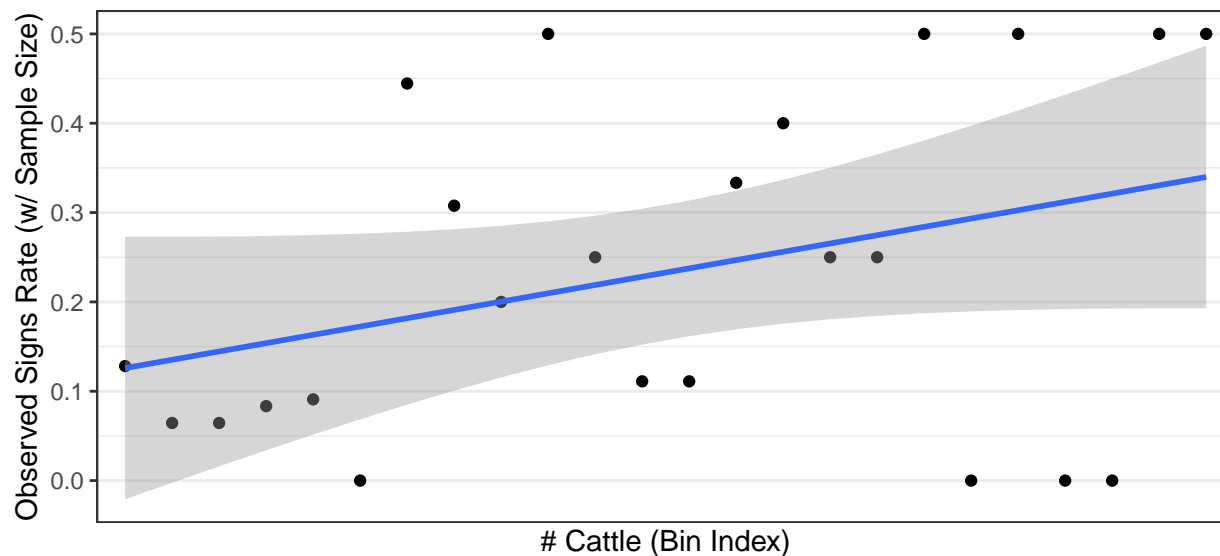
We also have several indicator variables to consider. To investigate these, we can look at the accuracy in predicting the response using each indicator as a sole predictor.

Indicator	Accuracy
accessible_cattle_house_present	0.3590
accessible_feed_present	**0.2491**
grass_silage	0.6154
cereal_silage	0.6740
hay_straw	0.3516
cereal_grains	0.5348
concentrates	0.5568
proteinblocks	**0.8022**
sugarbeet	**0.7656**
vegetables	**0.7802**
molasses	0.6703

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 219  43
##          1  11   0
##
##                Accuracy : 0.8022
##                  95% CI : (0.7499, 0.8478)
##     No Information Rate : 0.8425
##     P-Value [Acc > NIR] : 0.969
##
##                   Kappa : -0.0686
##
##  Mcnemar's Test P-Value : 2.459e-05
##
##             Sensitivity : 0.9522
##             Specificity : 0.0000
##          Pos Pred Value : 0.8359
```

```
##           Neg Pred Value : 0.0000
##                Prevalence : 0.8425
##            Detection Rate : 0.8022
##      Detection Prevalence : 0.9597
##          Balanced Accuracy : 0.4761
##
##           'Positive' Class : 0
##
```

Note: accuracy is the proportion of the data that are predicted correctly. The accuracy is good for some indicators because most observations are 0 for the response. The confusion matrix shows that the indicators do not have as much predictive power as it may seem from the accuracy alone.

Finally, we may look at the signs rate for each year and season combination.
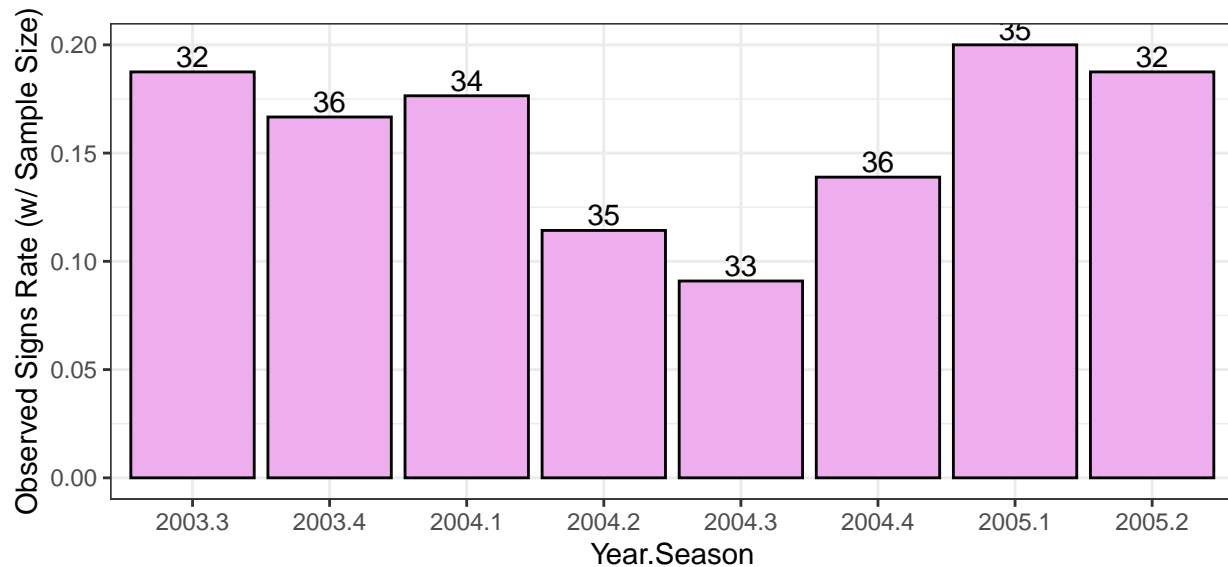
2003 and 2005 seem to have similar rates, but 2004 may have an appreciably lower rate of badger sign observations.

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## signs_in_yard ~ no_active_setts_in_fields + no_cattle_in_buildings_yard +
##     year + (1 | farm_code_numeric)
##    Data: model_data
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##    158.6    180.2    -73.3    146.6      267
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9952 -0.1948 -0.1043 -0.0618  5.8024
##
## Random effects:
##  Groups            Name           Variance Std.Dev.
```

4

```
##   farm_code_numeric (Intercept) 6.514     2.552
## Number of obs: 273, groups:  farm_code_numeric, 36
##
## Fixed effects:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -5.102489   1.102378  -4.629 3.68e-06 ***
## no_active_setts_in_fields  0.510660   0.161161   3.169  0.00153 **
## no_cattle_in_buildings_yard 0.004465   0.003318   1.346  0.17844
## year2004                  -0.875619   0.636503  -1.376  0.16892
## year2005                   0.243728   0.686189   0.355  0.72245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) n_ctv___ n_ctt___ yr2004
## n_ctv_stt__ -0.598
## n_cttl_n_b_ -0.285 -0.035
## year2004    -0.215 -0.055   -0.076
## year2005    -0.369  0.133   -0.103    0.549


## $farm_code_numeric
```

farm_code_numeric