

# Stage 2: Exploratory Data Analysis

Isaac Slagel and Jack Welsh

4/18/2019

## Main Report

Table 1: Modeling Dog Outcomes in Dallas Animal Shelters

	<i>Dependent variable:</i>		
	Proportion of dogs who died		
	(1)	(2)	(3)
Pitbull	0.116 (-2.061,2.293)	0.111 (-2.046,2.267)	0.552 (-1.577,2.680)
Scannable Chip	3.440*** (1.192,5.688)	3.424*** (1.206,5.642)	3.489*** (1.405,5.572)
Summer Outcome	0.789 (-1.488,3.067)	0.799 (-1.445,3.043)	0.781 (-1.314,2.875)
Contagious	1.461 (-0.805,3.726)	1.447 (-0.786,3.681)	1.478 (-0.611,3.568)
Treatable at intake		7.286*** (4.485,10.086)	3.975*** (1.615,6.336)
Intercept			0.161 (-1.967,2.290)
Observations	195	195	195

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Annotated Appendix

### Pitbulls

One of our research interests was to look into how different species of dogs fare in the animal shelter system. Specifically, we were interested in why pitbulls are more likely to be dead at the outcome of their time in the shelter. Does this trend continue after controlling for other variables in our dataset?

Initially we fit a model which factors in the chip status, whether or not the outcome was in the summer, and whether the dog was a pitbull or not.

```
pitbull_binom <- adoptions %>%  
  filter(!str_detect(intake_subtype, "(DEAD)|(DIED)")) %>%  
  filter(dog == 1) %>%  
  mutate(out_dead = outcome_type %in% c("DEAD ON ARRIVAL", "EUTHANIZED", "DIED"),  
         summer = ifelse(month %in% c(5, 6, 7, 8, 9), 1, 0)) %>%
```

```

mutate(chip_status = ifelse(chip_status=="SCAN CHIP", 1, 0)) %>%
group_by(pitbull, chip_status, summer, month) %>%
summarize(prop_dead = sum(out_dead)/n(), count = n())

pitbull_model1_binom <- glm(prop_dead ~ pitbull + chip_status + summer, weights = count, family = binomial)

pitbull_model1_quasi <- glm(prop_dead ~ pitbull + chip_status + summer, weights = count, family = quasibinomial)

summary(pitbull_model1_binom)

##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer, family = binomial,
##      data = pitbull_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -7.4184  -1.8174  -0.4575   1.5859   9.4807
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.15259    0.02131 -100.994 < 2e-16 ***
## pitbull      1.23537    0.02783  44.387 < 2e-16 ***
## chip_status -0.23649    0.03050  -7.753 8.97e-15 ***
## summer       0.37879    0.02938  12.891 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2641.29  on 51  degrees of freedom
## Residual deviance:  544.19  on 48  degrees of freedom
##      (2 observations deleted due to missingness)
## AIC: 867.48
##
## Number of Fisher Scoring iterations: 4
exp(confint(pitbull_model1_binom))

## Waiting for profiling to be done...
##              2.5 %    97.5 %
## (Intercept) 0.1114096 0.1211179
## pitbull     3.2569778 3.6324244
## chip_status 0.7434426 0.8378715
## summer      1.3786281 1.5469365
summary(pitbull_model1_quasi)

##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer, family = quasibinomial,
##      data = pitbull_binom, weights = count)
##
## Deviance Residuals:

```

```
##      Min      1Q   Median      3Q      Max
## -7.4184 -1.8174 -0.4575   1.5859   9.4807
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.15259    0.06971 -30.880 < 2e-16 ***
## pitbull      1.23537    0.09103  13.572 < 2e-16 ***
## chip_status -0.23649    0.09976  -2.371 0.021826 *
## summer       0.37879    0.09610   3.942 0.000262 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 10.69677)
##
## Null deviance: 2641.29 on 51 degrees of freedom
## Residual deviance: 544.19 on 48 degrees of freedom
## (2 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
exp(confint(pitbull_model1_quasi))
```

```
## Waiting for profiling to be done...
```

```
##              2.5 %    97.5 %
## (Intercept) 0.1011467 0.1329404
## pitbull      2.8769183 4.1109439
## chip_status  0.6478111 0.9580232
## summer       1.2082342 1.7612914
```

Initially we see that all variables included have a significant relationship to the proportion of dogs who are dead at the end of their time in the shelter, even after inflating our standard errors to account for the variance structure of our data.

Next lets try to control for whether or not a animal was contagious when it was brought into the shelter.

```
pitbull_binom <- adoptions %>%
  filter(!str_detect(intake_subtype, "(DEAD)|(DIED)")) %>%
  filter(dog == 1) %>%
  mutate(out_dead = outcome_type %in% c("DEAD ON ARRIVAL", "EUTHANIZED", "DIED"),
         summer = ifelse(month %in% c(5, 6, 7, 8, 9), 1, 0),
         chip_status = ifelse(chip_status=="SCAN CHIP", 1, 0),
         contagious=ifelse(grepl(".*[~NON-]CONTAGIOUS", intake_condition),1,0)) %>%
  group_by(pitbull, chip_status, summer, month, contagious) %>%
  summarize(prop_dead = sum(out_dead)/n(), count = n())

pitbull_model2_binom <- glm(prop_dead ~ pitbull + chip_status +
  summer + contagious, weights = count,
  family = binomial, data = pitbull_binom)

pitbull_model2_quasi <- glm(prop_dead ~ pitbull + chip_status +
  summer + contagious, weights = count,
  family = quasibinomial, data = pitbull_binom)

summary(pitbull_model2_binom)
```

```
##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer + contagious,
##      family = binomial, data = pitbull_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -7.204  -1.324  -0.415   1.069   9.546
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.20165    0.02169 -101.518 < 2e-16 ***
## pitbull      1.23078    0.02814  43.742 < 2e-16 ***
## chip_status -0.22483    0.03077  -7.306 2.74e-13 ***
## summer       0.36978    0.02969  12.456 < 2e-16 ***
## contagious   1.98593    0.08113  24.477 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 3266.27  on 100  degrees of freedom
## Residual deviance:  599.14  on  96  degrees of freedom
## (2 observations deleted due to missingness)
## AIC: 1045.2
##
## Number of Fisher Scoring iterations: 4
exp(confint(pitbull_model2_binom))

## Waiting for profiling to be done...

##              2.5 %    97.5 %
## (Intercept) 0.1059964 0.1154017
## pitbull      3.2401347 3.6179798
## chip_status  0.7517627 0.8481433
## summer       1.3654369 1.5339545
## contagious   6.2166833 8.5454709
summary(pitbull_model2_quasi)

##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer + contagious,
##      family = quasibinomial, data = pitbull_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -7.204  -1.324  -0.415   1.069   9.546
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.20165    0.05297 -41.566 < 2e-16 ***
## pitbull      1.23078    0.06872  17.910 < 2e-16 ***
## chip_status -0.22483    0.07516  -2.992 0.00353 **
## summer       0.36978    0.07251   5.100 1.71e-06 ***
```

```
## contagious    1.98593    0.19816  10.022  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 5.965136)
##
## Null deviance: 3266.27  on 100  degrees of freedom
## Residual deviance:  599.14  on  96  degrees of freedom
## (2 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
exp(confint(pitbull_model2_quasi))
```

```
## Waiting for profiling to be done...
```

```
##              2.5 %      97.5 %
## (Intercept) 0.09959825 0.1225854
## pitbull     2.99206819 3.9173028
## chip_status 0.68845298 0.9243987
## summer      1.25477039 1.6673621
## contagious  4.94800021 10.7784473
```

Next lets try to control for whether or not a animal was treatable when it was brought into the shelter.

```
pitbull_binom <- adoptions %>%
  filter(!str_detect(intake_subtype, "(DEAD)|(DIED)")) %>%
  filter(dog == 1) %>%
  mutate(out_dead = outcome_type %in% c("DEAD ON ARRIVAL", "EUTHANIZED", "DIED"),
         summer = ifelse(month %in% c(5, 6, 7, 8, 9), 1, 0),
         chip_status = ifelse(chip_status=="SCAN CHIP", 1, 0),
         contagious=ifelse(grepl("[^NON-]CONTAGIOUS", intake_condition),1,0),
         treatable=ifelse(grepl("^TREATABLE.*", intake_condition),1,0)) %>%
  group_by(pitbull, chip_status, summer, month, contagious, treatable) %>%
  summarize(prop_dead = sum(out_dead)/n(), count = n())

pitbull_model3_binom <- glm(prop_dead ~ pitbull + chip_status +
  summer + contagious + treatable, weights = count,
  family = binomial, data = pitbull_binom)

pitbull_model3_quasi <- glm(prop_dead ~ pitbull + chip_status +
  summer + contagious + treatable , weights = count,
  family = quasibinomial, data = pitbull_binom)

summary(pitbull_model3_binom)
```

```
##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer + contagious +
##       treatable, family = binomial, data = pitbull_binom, weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6.0491  -1.4236  -0.1897   0.9336   9.5936
##
```

```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.59457    0.03949  -15.06 < 2e-16 ***
## pitbull      1.24955    0.02925   42.72 < 2e-16 ***
## chip_status  -0.24774    0.03176   -7.80 6.21e-15 ***
## summer        0.39095    0.03070   12.73 < 2e-16 ***
## contagious    1.38010    0.08913   15.48 < 2e-16 ***
## treatable    -1.82378    0.03953  -46.14 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 5475.42 on 194 degrees of freedom
## Residual deviance: 842.76 on 189 degrees of freedom
## (2 observations deleted due to missingness)
## AIC: 1514.1
##
## Number of Fisher Scoring iterations: 4
exp(confint(pitbull_model3_binom))

## Waiting for profiling to be done...

##           2.5 %    97.5 %
## (Intercept) 0.5106406 0.5961339
## pitbull      3.2943807 3.6946299
## chip_status  0.7333041 0.8305406
## summer       1.3918868 1.5699049
## contagious   3.3384762 4.7351594
## treatable    0.1493824 0.1744212
summary(pitbull_model3_quasi)

##
## Call:
## glm(formula = prop_dead ~ pitbull + chip_status + summer + contagious +
##      treatable, family = quasibinomial, data = pitbull_binom,
##      weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6.0491  -1.4236  -0.1897   0.9336   9.5936
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.59457    0.08240  -7.216 1.26e-11 ***
## pitbull      1.24955    0.06104  20.472 < 2e-16 ***
## chip_status  -0.24774    0.06628  -3.738 0.000246 ***
## summer        0.39095    0.06407   6.102 5.80e-09 ***
## contagious    1.38010    0.18600   7.420 3.86e-12 ***
## treatable    -1.82378    0.08249 -22.110 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 4.354402)
```

```
##
## Null deviance: 5475.42 on 194 degrees of freedom
## Residual deviance: 842.76 on 189 degrees of freedom
## (2 observations deleted due to missingness)
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
exp(confint(pitbull_model3_quasi))
```

```
## Waiting for profiling to be done...
```

```
##           2.5 %    97.5 %
## (Intercept) 0.4692407 0.6482124
## pitbull     3.0952904 3.9321196
## chip_status 0.6848701 0.8881191
## summer      1.3032732 1.6754195
## contagious  2.7616713 5.7307395
## treatable   0.1373228 0.1897619
```

## Dogs vs Cats

```
cat_dog_binom <- adoptions %>%
  filter(!str_detect(intake_subtype, "(DEAD)|(DIED)")) %>%
  filter(dog == 1 | cat == 1) %>%
  mutate(out_dead = outcome_type %in% c("DEAD ON ARRIVAL", "EUTHANIZED", "DIED"),
         summer = ifelse(month %in% c(5, 6, 7, 8, 9), 1, 0)) %>%
  mutate(chip_status = ifelse(chip_status=="SCAN CHIP", 1, 0)) %>%
  group_by(dog, chip_status, summer, stray) %>%
  summarize(prop_dead = sum(out_dead)/n(), count = n())
```

```
dogcat_model1_binom <- glm(prop_dead ~ dog, weights = count, family = binomial, data = cat_dog_binom)
```

```
dogcat_model1_quasi <- glm(prop_dead ~ dog, weights = count, family = quasibinomial, data = cat_dog_binom)
```

```
summary(dogcat_model1_binom)
```

```
##
## Call:
## glm(formula = prop_dead ~ dog, family = binomial, data = cat_dog_binom,
##      weights = count)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -19.837   -7.385   -2.250    4.985   15.639
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.43741    0.02199  -65.38  <2e-16 ***
## dog          -0.30509    0.02563  -11.90  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 1883.0 on 17 degrees of freedom
## Residual deviance: 1745.6 on 16 degrees of freedom
## AIC: 1866
##
## Number of Fisher Scoring iterations: 4
exp(confint(dogcat_model1_binom))

## Waiting for profiling to be done...
##
## 2.5 % 97.5 %
## (Intercept) 0.2274800 0.2479563
## dog 0.7010303 0.7751213
summary(dogcat_model1_quasi)

##
## Call:
## glm(formula = prop_dead ~ dog, family = quasibinomial, data = cat_dog_binom,
## weights = count)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -19.837 -7.385 -2.250 4.985 15.639
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.4374 0.2249 -6.393 8.91e-06 ***
## dog -0.3051 0.2621 -1.164 0.261
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 104.5899)
##
## Null deviance: 1883.0 on 17 degrees of freedom
## Residual deviance: 1745.6 on 16 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
exp(confint(dogcat_model1_quasi))

## Waiting for profiling to be done...
##
## 2.5 % 97.5 %
## (Intercept) 0.1495373 0.3624932
## dog 0.4457027 1.2501118
dogcat_model2_binom <- glm(prop_dead ~ summer, weights = count, family = binomial, data = cat_dog_binom)
dogcat_model2_quasi <- glm(prop_dead ~ summer, weights = count, family = quasibinomial, data = cat_dog_binom)
summary(dogcat_model2_binom)

##
## Call:
## glm(formula = prop_dead ~ summer, family = binomial, data = cat_dog_binom,
```



```

##      weights = count)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -18.202   -7.386   -4.440    5.671   17.663
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.81259    0.01405  -129.0  <2e-16 ***
## summer      0.45539    0.02372   19.2  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1883.0  on 17  degrees of freedom
## Residual deviance: 1524.9  on 16  degrees of freedom
## AIC: 1645.3
##
## Number of Fisher Scoring iterations: 4
exp(confint(dogcat_model2_binom))

## Waiting for profiling to be done...

##              2.5 %    97.5 %
## (Intercept) 0.1587829 0.1677739
## summer      1.5050770 1.6517349
summary(dogcat_model2_quasi)

##
## Call:
## glm(formula = prop_dead ~ summer, family = quasibinomial, data = cat_dog_binom,
##      weights = count)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -18.202   -7.386   -4.440    5.671   17.663
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.8126    0.1344 -13.490 3.71e-10 ***
## summer       0.4554    0.2268  2.008  0.0619 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 91.44736)
##
##      Null deviance: 1883.0  on 17  degrees of freedom
## Residual deviance: 1524.9  on 16  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4

```

```
exp(confint(dogcat_model2_quasi))
```

```
## Waiting for profiling to be done...
```

```
##              2.5 %    97.5 %
```

```
## (Intercept) 0.1243408 0.2107409
```

```
## summer      1.0043970 2.4491693
```

**Multilevel on Council District**