

Hw6_q3

10/7/2019

1.

Myopathy is columns, and cows outcome is on teh y. We see that the percent of surviving cow (outcome =1) is 39% for those without myopathy and 6% for those with myopathy.

```
##
##           0           1
##  0 0.61417323 0.93684211
##  1 0.38582677 0.06315789
```

2. Fit logistic regression.

Based on the coefficient table the estimated log probability when myopathy = 0 would be -0.4649 ± 0.1823 , and when myopathy = 1 would be -2.2320 ± 0.4595 . The decrease in odds of survival from myopathy is then $\exp(-2.2320)=0.11$. Then the probability of survival when myopathy=1 is $0.11*\exp(-0.4649) = 0.07$, and when myopathy=0 is $1-\exp(-0.4649) = 0.37$. These numbers are about where the last row from Q1 are.

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##      select

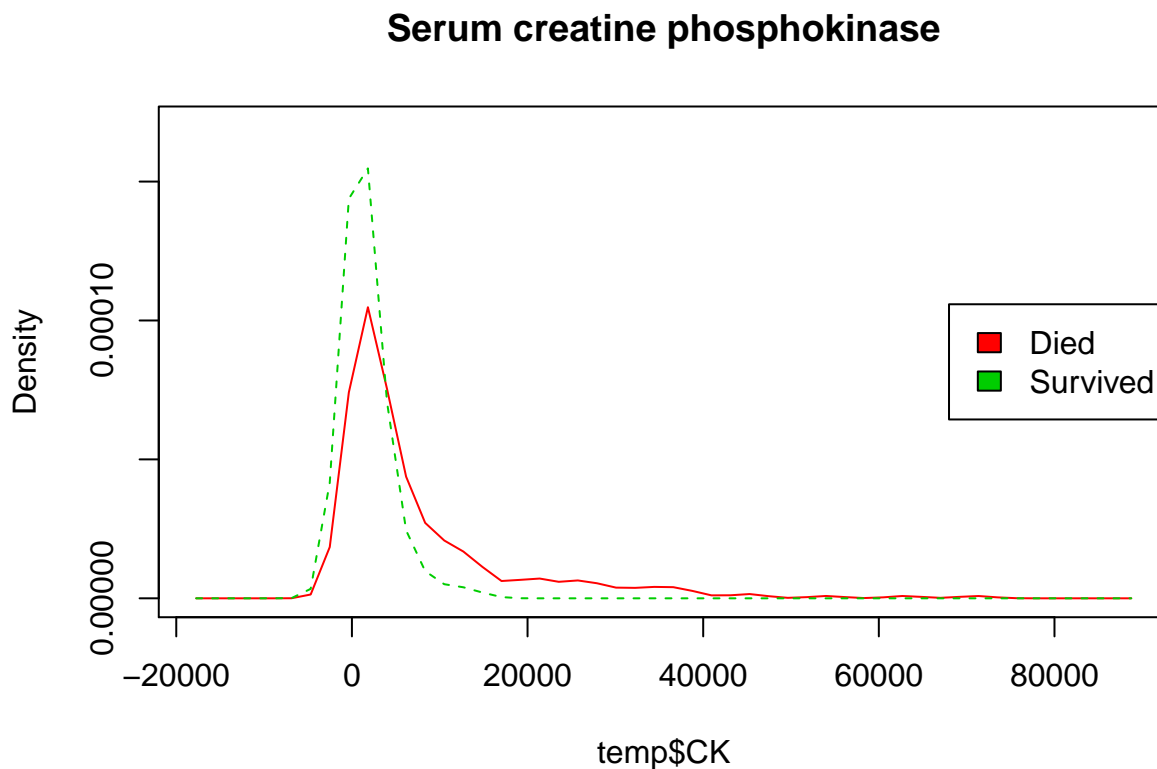
## The following object is masked from 'package:alr3':
##
##      forbes

##
## Call:
## glm(formula = Outcome ~ Myopathy, family = binomial(link = "logit"),
##      data = downer)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9874  -0.9874  -0.3612  -0.3612   2.3504
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.4649     0.1823  -2.550   0.0108 *
## Myopathy      -2.2320     0.4595  -4.858 1.19e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 248.57  on 221  degrees of freedom
## Residual deviance: 214.14  on 220  degrees of freedom
## (213 observations deleted due to missingness)
## AIC: 218.14
##
## Number of Fisher Scoring iterations: 5
```

3.

We plotted the density of the CK and log(CK) below. It seems like the dead cows had higher levels of phosphokinase, and this is a bit clearer to see when we have log scale x-axis. Also the values look more normally distributed in the second graph.

```
## Package 'sm', version 2.2-5.6: type help(sm) for summary information
##
## Attaching package: 'sm'
## The following object is masked from 'package:MASS':
##
## muscle
```



4

Here are the results, showing that the fold increase in survival for each unit increase in log(CK) is $\exp(-0.6117) = 0.542428$. This suggests perhaps myopathy is more predictive.

```
##
## Call:
## glm(formula = Outcome ~ log(CK), family = binomial(link = "logit"),
## data = downer)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1337  -0.8811  -0.5608   1.0588   1.9935
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    4.0007     0.5809   6.887 5.69e-12 ***
```

```
## log(CK)      -0.6117      0.0793  -7.714 1.22e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 550.49  on 412  degrees of freedom
## Residual deviance: 475.18  on 411  degrees of freedom
## (22 observations deleted due to missingness)
## AIC: 479.18
##
## Number of Fisher Scoring iterations: 3
```

5

Here are the results of the model:

```
##
## Call:
## glm(formula = Outcome ~ log(CK) + Myopathy + log(CK):Myopathy,
##      family = binomial(link = "logit"), data = downer)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2221  -0.9671  -0.3403   0.9456   2.4952
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.02810    1.16564   0.024  0.9808
## log(CK)        -0.06788    0.15872  -0.428  0.6689
## Myopathy        5.31297    3.84652   1.381  0.1672
## log(CK):Myopathy -0.81290    0.44494  -1.827  0.0677 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 246.27  on 217  degrees of freedom
## Residual deviance: 208.86  on 214  degrees of freedom
## (217 observations deleted due to missingness)
## AIC: 216.86
##
## Number of Fisher Scoring iterations: 6
```

Additionally we can show the deviance per term:

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Outcome
##
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
```

```
## NULL                217      246.27
## log(CK)             1  21.3755      216      224.90 3.776e-06 ***
## Myopathy            1  12.6153      215      212.28 0.0003826 ***
## log(CK):Myopathy    1   3.4204      214      208.86 0.0643943 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It seems like the effect of each of log(CK) and Myopathy is significant ($p < 0.001$) but their interaction is not.