Swetha Adike

Venu Goud Raparti

Problems from Chapter 20

MSIS 545

9.

This problem is about calculating log odds of survival in Donner party case study. The case study is about survival study when the Donner and Reed families left Springfield, Illinois for California in 1846. When the Donner Party reached For Bridger in July, the leaders decided to attempt a new untested route to Sacramento Valley with the full size of 87 people and 20 wagons. The group became stranded in eastern Sierra Nevada mountains as the region was hit by heavy snow and by the time the last survivor was rescued 40 of 87 members had died from famine and exposure to extreme cold. The data contains the ages and sexes of the adult survivors and nonsurvivors of the party over the years of the journey. This was taken up by the anthropologists to study if females are better able to withstand harsh weather conditions than males.

In connection to the collected data, the log odds of survival for females is estimated to be 3.2 – (0.078 * age) and 1.6 – (0.078 * age) for males.

a. Let's estimate the probabilities for survival for men and women of 25- and 50-years age

The probability of survival for female of 25 years age is

```
> f25 <- 3.2 - (0.078 * 25)
> f25
[1] 1.25
> fsp25 <- exp(f25)/(1+exp(f25))
> fsp25
[1] 0.7772999
```

The probability of survival for female of 50 years age is

```
> f50 <- 3.2 - (0.078 * 50)
> f50
[1] -0.7
> fsp50 <- exp(f50)/(1+exp(f50))
> fsp50
```

```
[1] 0.3318122
```

The probability of survival for male of 25 years age is

```
> m25 <- 1.6 - (0.078 * 25)
> m25
[1] -0.35
> msp25 <- exp(m25)/(1+exp(m25))
> msp25
[1] 0.4133824
```

The probability of survival for male of 50 years age is

```
> m50 <- 1.6 - (0.078 * 50)
> m50
[1] -2.3
> msp50 <- exp(m50)/(1+exp(m50))
> msp50
[1] 0.09112296
```

b. Given the estimated probability of survival is 0.5

=> logit(0.5) = log(0.5/(1-0.5)) = 0

From the given log odds survival equation for female, age = 3.2 – 0/0.078 = 41.02 ~ 41 years

From the given log odds survival equation for male, age = 1.6 – 0/0.078 = 20.51 ~ 20 years

---

12.

This problem is regarding Duchenne muscular dystrophy (DMD) which shows no symptoms in women and genetically transmitted from mother to child. As the female with this disease show no symptoms, they may unknowingly carry to their offspring. And so doctors had to rely on some kind of test to detect the presence of disease.

The data contains the levels of two enzymes in blood – Creating Kinase (CK) and Hemopexin (H) for 38 known DMS carriers and 82 women who are not carriers. Using this data let's find if there is any relevance of these enzymes on the disease and find an equation for indicating whether a woman is a likely carrier.

Let's look into the summary of the data set

```
> summary(ex2012)
```

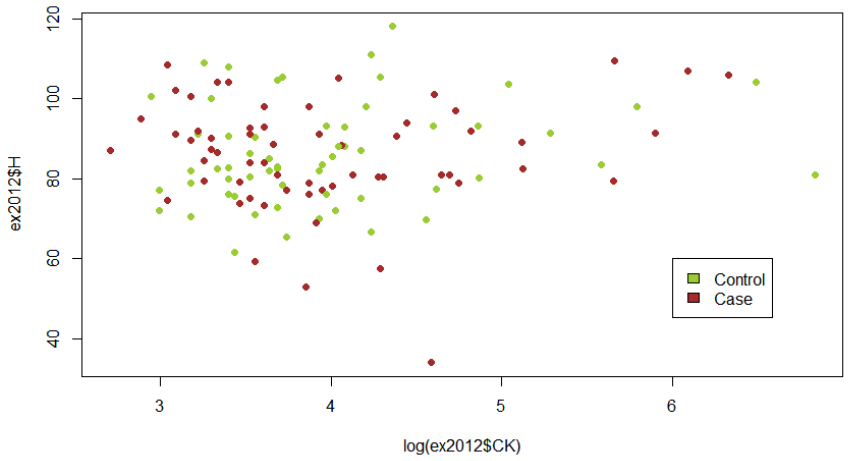
```
     Group          CK                H
 Case   :38   Min.   : 15.00   Min.   : 34.00
 Control:82   1st Qu.: 30.00   1st Qu.: 78.72
              Median : 41.50   Median : 85.25
              Mean   : 83.35   Mean   : 86.23
              3rd Qu.: 73.00   3rd Qu.: 93.40
              Max.   :925.00   Max.   :118.00
```

The summary shows that the CK field should be transformed as the scatter is more

a. Let's draw a scatterplot of H versus log(CK)

```
> plot(log(ex2012$CK), ex2012$H, pch = 19, col = rep(c('yellowgreen',
'brown'), each = 2), cex = 1)
```

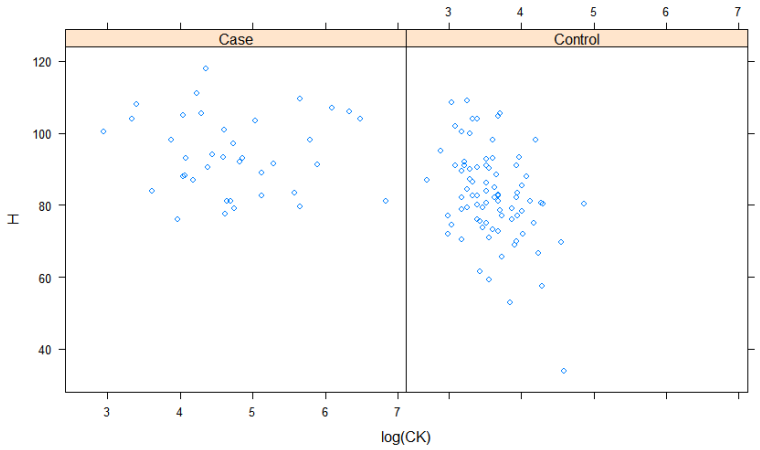
```
> legend(6,60, legend = c("Control", "Case"), fill =
c("yellowgreen","brown"))
```



The above shows no relationship between log(CK) and hemopexin though most of the cases are located above 75 Hemopexin

The below code is for another plot to represent controls and carriers

```
> xyplot(H ~ log(CK) | Group, data = ex2012 )
```



The above plot shows little separation between carriers and controls

b. Let's fit a logistic regression for carrier on CK and CK square

```
> carrier.glm <- glm( Group ~ CK + I(CK^2), data = ex2012,
family = "binomial")

> summary(carrier.glm)

Call:

glm(formula = Group ~ CK + I(CK^2), family = "binomial", data =
ex2012)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-2.50536  -0.03915   0.37969   0.51841   2.27337

Coefficients:

              Estimate Std. Error z value Pr(>|z|)

(Intercept)  4.177e+00  7.264e-01   5.751 8.87e-09 ***

CK          -5.798e-02  1.299e-02  -4.463 8.10e-06 ***

I(CK^2)      5.054e-05  3.268e-05   1.547    0.122

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 149.84  on 119  degrees of freedom

Residual deviance:  85.47  on 117  degrees of freedom
```

```
AIC: 91.47

Number of Fisher Scoring iterations: 9
```

The above coefficient for CK square shows that it is not significant with p value 0.122

Now let's fit logistic regression for carrier on log(CK) and log(CK square)

```
> lcarrier.glm <- glm( Group ~ log(CK) + I(log(CK)^2), data =
ex2012, family = "binomial")

> summary(lcarrier.glm)

Call:

glm(formula = Group ~ log(CK) + I(log(CK)^2), family =
"binomial",

    data = ex2012)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-2.39368  -0.03111   0.38041   0.50222   2.28558

Coefficients:

             Estimate Std. Error z value Pr(>|z|)

(Intercept)    -9.735     16.298  -0.597    0.550

log(CK)         8.516      8.358   1.019    0.308

I(log(CK)^2)   -1.446      1.063  -1.360    0.174

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 149.840  on 119  degrees of freedom

Residual deviance:  85.017  on 117  degrees of freedom

AIC: 91.017

Number of Fisher Scoring iterations: 7
```

Here too the log(CK) square term is not significant with p value 0.174

Comparing both the above models, CK for untransformed model seems more appropriate since the p value is <0.000001 for this and coefficient of log(CK) is not significant in the other model

c. Let's fit a logistic regression model for carrier on log(CK) and H. Below is the code for this

```
> carrier2.glm <- glm( Group ~ log(CK) + H, data = ex2012,
family = "binomial")
```

```
> summary(carrier2.glm)

Call:

glm(formula = Group ~ log(CK) + H, family = "binomial", data =
ex2012)

Deviance Residuals:

     Min        1Q    Median        3Q       Max

-2.60371  -0.09903   0.16696   0.38782   1.89706


Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept) 28.91340    5.80017   4.985 6.20e-07 ***

log(CK)     -4.02043    0.82910  -4.849 1.24e-06 ***

H           -0.13652    0.03654  -3.736 0.000187 ***

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 149.840  on 119  degrees of freedom

Residual deviance:  61.992  on 117  degrees of freedom

AIC: 67.992

Number of Fisher Scoring iterations: 7
```

The above model seems to be more significant as all the coefficients are significant with p value <0.0001. Below are the coefficients

```
> carrier2.glm$coefficients

(Intercept)     log(CK)           H

 28.9134030  -4.0204252  -0.1365189

And the standard errors are

5.80017        0.82910   0.03654
```

d. Let’s fit logistic regression model without considering the enzyme fields

```
> carrier3.glm <- glm( Group ~1, data = ex2012, family =
"binomial")
```

```
> summary(carrier3.glm)

Call:

glm(formula = Group ~ 1, family = "binomial", data = ex2012)

Deviance Residuals:

    Min       1Q   Median       3Q      Max

-1.5165  -1.5165   0.8727   0.8727   0.8727

Coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)   0.7691     0.1962   3.919 8.88e-05 ***

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 149.84  on 119  degrees of freedom

Residual deviance: 149.84  on 119  degrees of freedom

AIC: 151.84

Number of Fisher Scoring iterations: 4
```

And find out drop in deviance considering the model with log(CK), H and the model without log(CK) and H

```
> anova( carrier2.glm, carrier3.glm, test = "Chisq")

Analysis of Deviance Table

Model 1: Group ~ log(CK) + H

Model 2: Group ~ 1

  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)

1       117     61.992

2       119    149.840 -2  -87.847 < 2.2e-16 ***

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

From the above anova model, the drop-in-deviance is 87.847 (149.840 – 61.992) and p value for this test is <0.00001 which tells that there is convincing evidence that the enzymes creatine kinase and hemopexin are significant in the determining DMD disease in woman.

e. Given typical values for CK and H are 80 and 85 and suspected carrier has 300 and 100 respectively.

```
> eq <- carrier2.glm$coefficients
> sus <- eq[1]+eq[2]*log(300)+eq[3]*100
> sus
(Intercept)
  -7.670117
> typ <- eq[1]+eq[2]*log(80)+eq[3]*85
> typ
(Intercept)
  -0.308313
> sus/typ
(Intercept)
   24.87769
```

From above, the odds of suspected carriers with CK 300 and H 100 is -7.67 and the odds of typical carriers with CK 80 and H 85 is 0.0308. Odds ratio is ~ 25 which means that odds of suspected carriers is almost 25 times more than odds of a typical carrier woman.

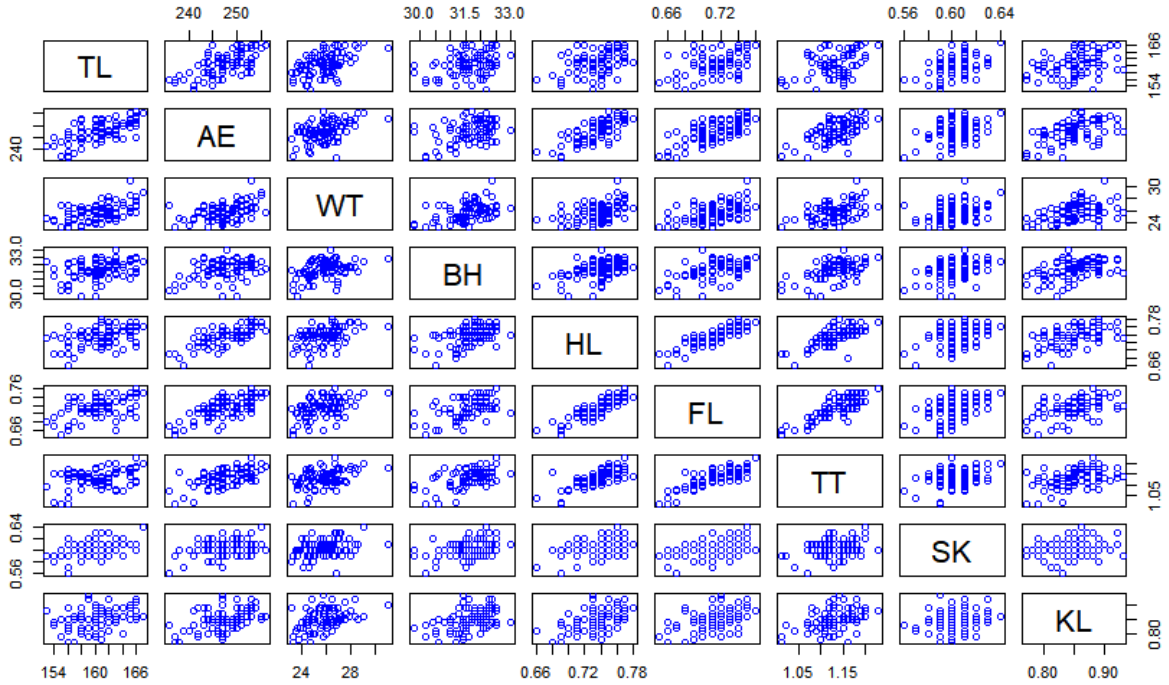
16.

This is a problem involving various characteristics of some house sparrows, perished and survived, which were found on the ground after a severe winter storm. The data set contains male sparrows with survival status SV (survived and perished), age AG (adult and juvenile), length from tip of beak to tip of tail (TL), length from tip to tip of extended wings (AE), weight (WT), length of head (BH), length of humerus arm bone (HL), length of femur thigh bone (FL), length of tibio-tarsus leg bone (TT), breadth of skull (SK), and length of sternum (KL).

Let’s make a analysis of this data to see whether the probability of survival is associated with physical characteristics of the birds.

Let’s take a look at the data

```
> plot(ex2016 [,-c(1,2)], col = "Blue")
```



There seems to be no linear no significant linear relationship between TL BH, SK and KL but other than these, all the other characteristics more or less follow linearity

Let's look at the summary statistics

```
> summary(ex2016)
```

```
      Status          AG              TL              AE
 Perished:36   Min.   :1.000   Min.   :153.0   Min.   :236.0
 Survived:51   1st Qu.:1.000   1st Qu.:158.0   1st Qu.:245.0
               Median :1.000   Median :160.0   Median :247.0
               Mean   :1.322   Mean   :160.4   Mean   :247.5
               3rd Qu.:2.000   3rd Qu.:162.5   3rd Qu.:251.0
               Max.   :2.000   Max.   :167.0   Max.   :256.0
WT              BH
Min.   :23.2    Min.   :29.80
1st Qu.:24.7    1st Qu.:31.40
Median :25.8    Median :31.70
Mean   :25.8    Mean   :31.64
```

```
3rd Qu.:26.7   3rd Qu.:32.10
Max.   :31.0   Max.   :33.00


       HL               FL               TT
 Min.   :0.6600   Min.   :0.6500   Min.   :1.010
 1st Qu.:0.7250   1st Qu.:0.7000   1st Qu.:1.110
 Median :0.7400   Median :0.7100   Median :1.130
 Mean   :0.7353   Mean   :0.7134   Mean   :1.131
 3rd Qu.:0.7500   3rd Qu.:0.7300   3rd Qu.:1.160
 Max.   :0.7800   Max.   :0.7600   Max.   :1.230
SK               KL
Min.   :0.5600   Min.   :0.7700
1st Qu.:0.5900   1st Qu.:0.8300
Median :0.6000   Median :0.8500
Mean   :0.6032   Mean   :0.8511
3rd Qu.:0.6100   3rd Qu.:0.8800
Max.   :0.6400   Max.   :0.9300
```

The median and mean of all the characteristics are near by to form any difference in significance

To find the significant characteristics to form whether the sparrow may perish or survive, let's make a backward propagation where all the characteristics are considered first and eliminated one by one which is of least or no significance.

So let's start the process by considering all the characteristics to fit a logistic regression model

```
> sparrows.glm <- glm(Status ~ AG + TL + AE+ WT + BH + HL + FL +
TT + SK + KL,
+                     family = binomial, data = ex2016)
> summary(sparrows.glm)
Call:
glm(formula = Status ~ AG + TL + AE + WT + BH + HL + FL + TT +
    SK + KL, family = binomial, data = ex2016)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2252  -0.5232   0.1397   0.5131   2.0134
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 27.45975   26.43983   1.039 0.299002
AG           0.10631    0.68253   0.156 0.876225
TL          -0.73634    0.18965  -3.883 0.000103 ***
AE           0.08275    0.12622   0.656 0.512060
WT          -0.88860    0.34182  -2.600 0.009333 **
BH           0.58293    0.59735   0.976 0.329131
HL          56.03494   31.05541   1.804 0.071176 .
FL          -6.64680   31.73442  -0.209 0.834096
TT           5.05213   14.05263   0.360 0.719210
SK          21.53121   27.28482   0.789 0.430037
KL          23.56111   12.03826   1.957 0.050326 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 118.01  on 86  degrees of freedom
Residual deviance:  65.92  on 76  degrees of freedom
AIC: 87.92
Number of Fisher Scoring iterations: 6
```

We can see that age coefficient has more p value 0.876, making it the first least significant factor. So removing this, let’s fit another logistic regression

```
> sparrows2.glm <- glm(Status ~ TL + AE+ WT + BH + HL + FL + TT
+ SK + KL,
+                     family = binomial, data = ex2016)
> summary(sparrows2.glm)
```

```
Call:
glm(formula = Status ~ TL + AE + WT + BH + HL + FL + TT + SK +
    KL, family = binomial, data = ex2016)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1979  -0.5187   0.1380   0.5195   1.9932
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  27.6154    26.5200   1.041  0.29773
TL           -0.7315     0.1865  -3.922 8.79e-05 ***
AE            0.0795     0.1246   0.638  0.52332
WT           -0.8930     0.3417  -2.613  0.00897 **
BH            0.5793     0.5973   0.970  0.33216
HL           56.3324    30.9729   1.819  0.06895 .
FL           -6.9118    31.6706  -0.218  0.82724
TT            5.1116    14.1011   0.362  0.71698
SK           21.7633    27.2708   0.798  0.42485
KL           23.5683    12.0734   1.952  0.05093 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 118.008  on 86  degrees of freedom
Residual deviance:  65.945  on 77  degrees of freedom
AIC: 85.945
Number of Fisher Scoring iterations: 6
```

From the above, FL coefficient has p value of 0.827, making it first least significant. So let’s fit another model removing this factor

```
> sparrows3.glm <- glm(Status ~ TL + AE+ WT + BH + HL + TT + SK
+ KL,
```

```
+                     family = binomial, data = ex2016)
> summary(sparrows3.glm)
Call:
glm(formula = Status ~ TL + AE + WT + BH + HL + TT + SK + KL,
    family = binomial, data = ex2016)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2211  -0.5397   0.1404   0.5014   1.9806
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 28.03611   26.36607   1.063  0.28763
TL          -0.73512    0.18643  -3.943 8.04e-05 ***
AE           0.07978    0.12459   0.640  0.52196
WT          -0.88694    0.33935  -2.614  0.00896 **
BH           0.55984    0.59108   0.947  0.34356
HL          52.95693   26.71797   1.982  0.04747 *
TT           3.39882   11.72971   0.290  0.77200
SK          22.04688   27.18725   0.811  0.41741
KL          23.41190   12.05676   1.942  0.05216 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 118.008  on 86  degrees of freedom
Residual deviance:  65.992  on 78  degrees of freedom
AIC: 83.992
Number of Fisher Scoring iterations: 6
```

We have to remove TT as its p value is 0.772 not anymore significant. Repeating the model again,

```
> sparrows4.glm <- glm(Status ~ TL + AE+ WT + BH + HL + SK + KL,
```

```
+                        family = binomial, data = ex2016)
> summary(sparrows4.glm)

Call:
glm(formula = Status ~ TL + AE + WT + BH + HL + SK + KL, family
= binomial,
    data = ex2016)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2095  -0.5401   0.1474   0.5000   2.0130

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  26.0639    25.4931   1.022  0.30660
TL           -0.7323     0.1854  -3.951 7.79e-05 ***
AE            0.0864     0.1226   0.705  0.48111
WT           -0.8791     0.3373  -2.607  0.00915 **
BH            0.5958     0.5761   1.034  0.30099
HL           56.0686    24.7957   2.261  0.02375 *
SK           22.2755    27.1914   0.819  0.41267
KL           23.3670    12.0203   1.944  0.05190 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 118.008  on 86  degrees of freedom
Residual deviance:  66.075  on 79  degrees of freedom
AIC: 82.075
Number of Fisher Scoring iterations: 6
```

From the above, AE coefficient is not significant with p value 0.481. Let’s iterate the model by removing this factor

```
> sparrows5.glm <- glm(Status ~ TL + WT + BH + HL + SK + KL,
```

```
+                      family = binomial, data = ex2016)
> summary(sparrows5.glm)
Call:
glm(formula = Status ~ TL + WT + BH + HL + SK + KL, family =
binomial,
    data = ex2016)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1559  -0.5221   0.1523   0.5308   1.9600
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  35.1445    21.8668   1.607  0.10801
TL           -0.6916     0.1744  -3.966 7.31e-05 ***
WT           -0.8473     0.3283  -2.581  0.00985 **
BH            0.5345     0.5576   0.959  0.33780
HL           65.1081    21.3385   3.051  0.00228 **
SK           20.9032    26.4328   0.791  0.42906
KL           24.6188    11.8527   2.077  0.03780 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 118.008  on 86  degrees of freedom
Residual deviance:  66.581  on 80  degrees of freedom
AIC: 80.581
Number of Fisher Scoring iterations: 6
```

P value for SK is 0.42, which is not anymore significant, let’s make the model without this factor

```
> sparrows6.glm <- glm(Status ~ TL + WT + BH + HL + KL,
+                      family = binomial, data = ex2016)
> summary(sparrows6.glm)
```

```
Call:
glm(formula = Status ~ TL + WT + BH + HL + KL, family =
binomial,
    data = ex2016)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9635  -0.5645   0.1492   0.6004   2.1646
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  41.5162    20.1743   2.058  0.03960 *
TL           -0.6888     0.1758  -3.919 8.91e-05 ***
WT           -0.8604     0.3240  -2.655  0.00792 **
BH            0.6371     0.5390   1.182  0.23716
HL           68.0605    20.9756   3.245  0.00118 **
KL           25.4462    11.9957   2.121  0.03390 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 118.008  on 86  degrees of freedom
Residual deviance:  67.214  on 81  degrees of freedom
AIC: 79.214
Number of Fisher Scoring iterations: 6
```

Here, other than BH all the other coefficients seem to be significant. Let’s remove this see what happens

```
> sparrows7.glm <- glm(Status ~ TL + WT + HL + KL,
+                      family = binomial, data = ex2016)
> summary(sparrows7.glm)
Call:
glm(formula = Status ~ TL + WT + HL + KL, family = binomial,
```

```
    data = ex2016)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2234  -0.5648   0.1540   0.6094   2.2701
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  49.9861    18.4879   2.704 0.006857 **
TL           -0.6573     0.1683  -3.907 9.35e-05 ***
WT           -0.7896     0.3097  -2.549 0.010800 *
HL           72.3327    20.7640   3.484 0.000495 ***
KL           27.3775    11.7780   2.324 0.020101 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 118.008  on 86  degrees of freedom
Residual deviance:  68.612  on 82  degrees of freedom
AIC: 78.612
Number of Fisher Scoring iterations: 6
```

All the coefficients are significant now with p value <0.05. So finally the characteristics TL, WT, HL and KL seem to be significant characteristics in determining whether the male sparrow has survived or perished.

In the above process, the residual deviance has increased a bit while dropping the factors. The difference is not much (65.92 when all the factors are considered to 68.612)

Let’s take a look at drop-in-deviance when no characteristic is involved

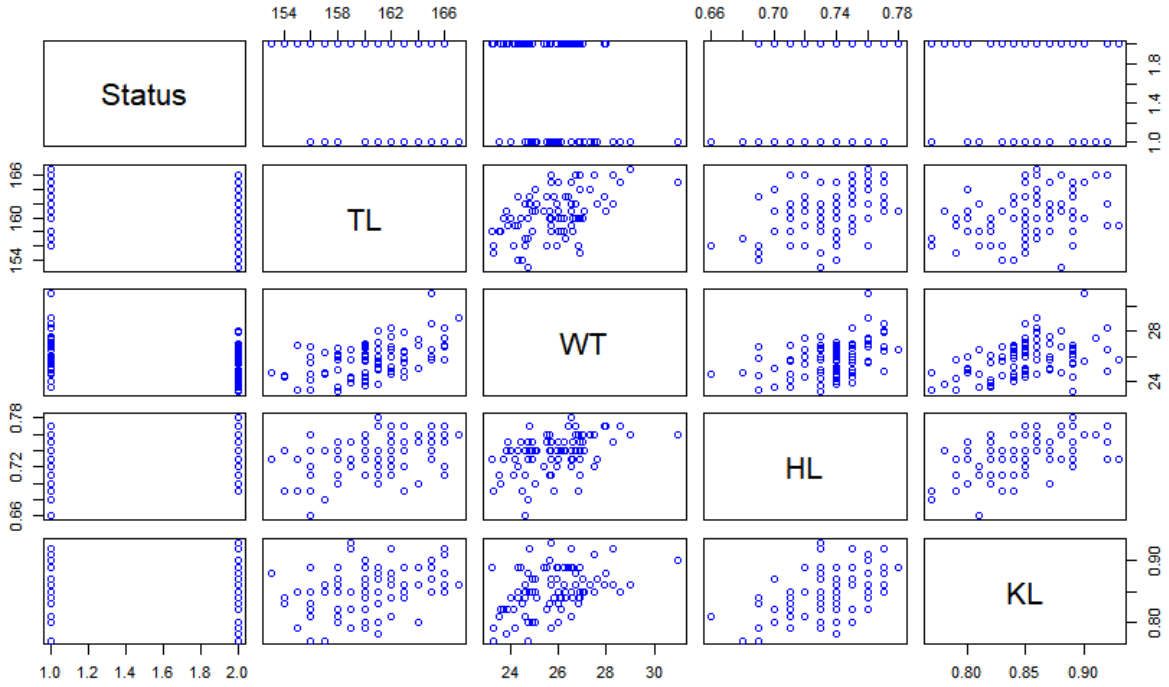
```
> sparrows8.glm <- glm(Status ~ 1,
+                      family = binomial, data = ex2016)
> summary(sparrows8.glm)
Call:
glm(formula = Status ~ 1, family = binomial, data = ex2016)
```

```
Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.328  -1.328   1.034   1.034   1.034
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.3483     0.2177     1.6     0.11
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 118.01  on 86  degrees of freedom
Residual deviance: 118.01  on 86  degrees of freedom
AIC: 120.01
Number of Fisher Scoring iterations: 4
> anova( sparrows7.glm, sparrows8.glm, test = "Chisq")
Analysis of Deviance Table
Model 1: Status ~ TL + WT + HL + KL
Model 2: Status ~ 1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1        82     68.612
2        86    118.008 -4  -49.396 4.826e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The above anova test shows that there is convincing evidence that the odds of survival are associated with TL, WT, HL and KL

Let’s make a plot of these characteristics

```
> plot(ex2016 [,-c(2,4,6,8,9,10)], col = "Blue")
```



The above plot shows that more or less there is a positive linear relation between all the other characteristics, when associated with weight.