

# 5291 hw7

Yijin Wang

11/6/2020

Categorize 'weight' as a binary variable, with WeightGroup = 1 (or Low), if weight < 120 mg, and 0, Otherwise.

1. Consider comparing Diet Levels 1 and 4 on Day 21.
  - a) Determine whether there is association between Diet and WeightGroup, using logistic regression, without adjusting for Birth Weight. Interpret what the estimated parameters denote.

```
data("ChickWeight")
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.2.1      v purrr   0.3.3
## v tibble  3.0.4      v dplyr   1.0.2
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

df<-ChickWeight%>%mutate(WeightGroup=ifelse(weight<120, 1, 0))
day21<-df%>%filter(Time==21)
day0<-df%>%filter(Time==0)
diet1_4<-day21%>%filter(Diet %in% c(1,4))
log1<-glm(WeightGroup~Diet, family = "binomial", data=diet1_4)
summary(log1)

##
## Call:
## glm(formula = WeightGroup ~ Diet, family = "binomial", data = diet1_4)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.64442  -0.64442  -0.64442  -0.00008   1.82974
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.4663      0.6405  -2.289   0.0221 *
## Diet4        -18.0997    3584.6710  -0.005   0.9960
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
## Null deviance: 18.346 on 24 degrees of freedom
## Residual deviance: 15.442 on 23 degrees of freedom
## AIC: 19.442
##
## Number of Fisher Scoring iterations: 18
```

The P-value is 0.9960 > 0.05, we fail to reject the null hypothesis and conclude that there is no association between Diet and Weightgroup.

b) Repeat (a) adjusting for Birth Weight. Interpret what the estimated parameters denote.

```
day21_adj<-left_join(day21, day0, by="Chick")%>%
  mutate(weight=weight.x-weight.y,
         base=weight.y,
         Time=Time.x,
         Diet=Diet.x,
         WeightGroup=ifelse(weight<120, 1, 0))%>%
  select(c(weight, base, Time, Chick, Diet, WeightGroup))

diet1_4_adj<-day21_adj%>%filter(Diet %in% c(1,4))
log2<-glm(WeightGroup ~ Diet, family = "binomial", data = diet1_4_adj)
summary(log2)
```

```
##
## Call:
## glm(formula = WeightGroup ~ Diet, family = "binomial", data = diet1_4_adj)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17741  -1.17741  -0.00008   1.17741   1.17741
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.736e-16  5.000e-01   0.000   1.000
## Diet4       -1.957e+01  3.585e+03  -0.005   0.996
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 31.343 on 24 degrees of freedom
## Residual deviance: 22.181 on 23 degrees of freedom
## AIC: 26.181
##
## Number of Fisher Scoring iterations: 18
```

The P-value is 0.996 > 0.05, we fail to reject the null hypothesis and conclude that there is no association between Diet and Weightgroup.

2. Repeat 1 for all 4 Diet Levels

```
#Without adjustment
log3<-glm(WeightGroup ~ Diet, family = "binomial", data=day21)
summary(log3)
```

```
##
## Call:
## glm(formula = WeightGroup ~ Diet, family = "binomial", data = day21)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.64442  -0.64442  -0.00008  -0.00008   2.14597
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.4663     0.6405  -2.289   0.0221 *
## Diet2        -0.7309     1.2334  -0.593   0.5535
## Diet3       -18.0997    3400.7176  -0.005   0.9958
## Diet4       -18.0997    3584.6710  -0.005   0.9960
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 26.996  on 44  degrees of freedom
## Residual deviance: 21.944  on 41  degrees of freedom
## AIC: 29.944
##
## Number of Fisher Scoring iterations: 18
```

*#With adjustment*

```
log4<-glm(WeightGroup ~ Diet, family = "binomial", data=day21_adj)
summary(log4)
```

```
##
## Call:
## glm(formula = WeightGroup ~ Diet, family = "binomial", data = day21_adj)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17741  -0.66805  -0.45904  -0.00013   2.14597
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -9.145e-16  5.000e-01   0.000   1.0000
## Diet2       -1.386e+00  9.354e-01  -1.482   0.1383
## Diet3       -2.197e+00  1.167e+00  -1.883   0.0597 .
## Diet4       -1.857e+01  2.174e+03  -0.009   0.9932
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 50.053  on 44  degrees of freedom
## Residual deviance: 38.690  on 41  degrees of freedom
## AIC: 46.69
##
## Number of Fisher Scoring iterations: 17
```

For both dataset, we see that p-value for all groups are greater than 0.05, thus concluding that there is no association between Diet and Weightgroup.

3. Repeat 1 using the L-1 regularized logistic regression

```

library(glmnet)

## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
## Loaded glmnet 3.0-2
##
#Without adjustment
diet<-day21%>%mutate(Diet=as.numeric(Diet))%>%filter(Diet %in% c(1,4))%>%
  select(Diet)%>%as.matrix()
weightgroup<-day21%>%filter(Diet %in% c(1,4))%>%
  select(WeightGroup)%>%as.matrix()
diet<-cbind(temp=1, diet)
cvfit<-cv.glmnet(x=diet, y=weightgroup, family="binomial")

## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

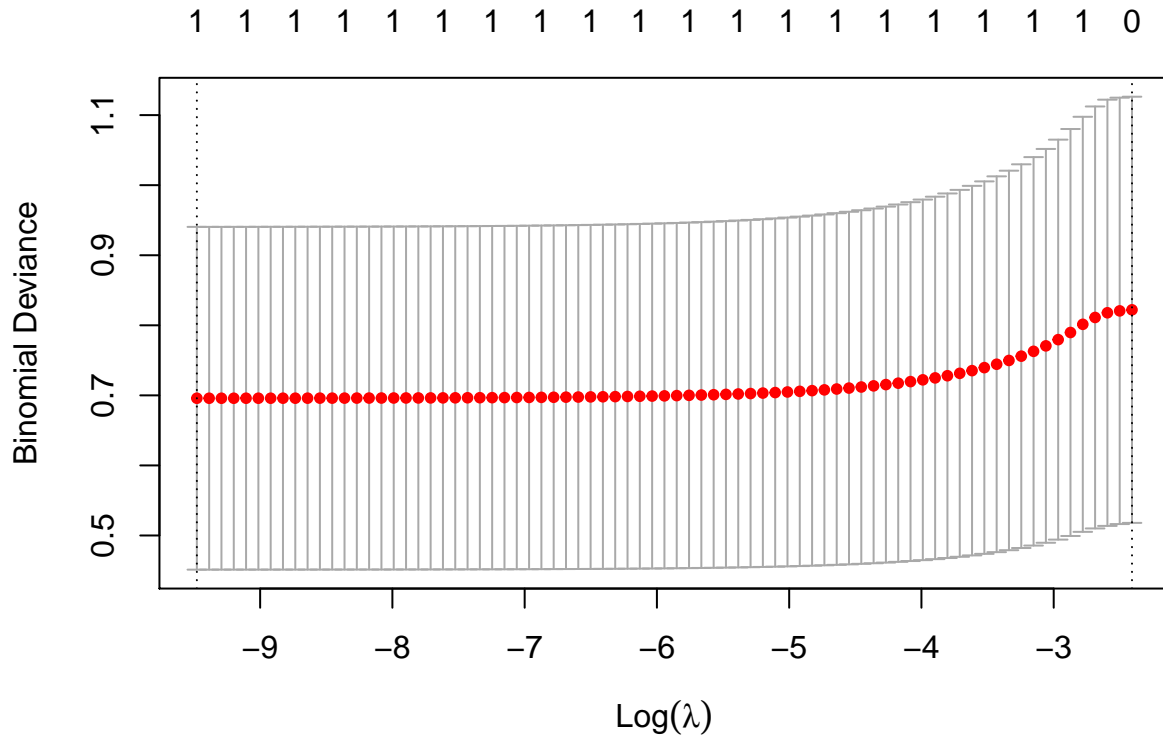
## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations per
## fold

```

```
plot(cvfit)
```



```
#With adjustment
```

```
diet1<-day21_adj%>%mutate(Diet=as.numeric(Diet))%>%filter(Diet %in% c(1,4))%>%
  select(Diet)%>%as.matrix()
weightgroup1<-day21_adj%>%filter(Diet %in% c(1,4))%>%
  select(WeightGroup)%>%as.matrix()
diet1<-cbind(temp=1, diet1)
cvfit1<-cv.glmnet(x=diet1, y=weightgroup1, family="binomial")
```

```
## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

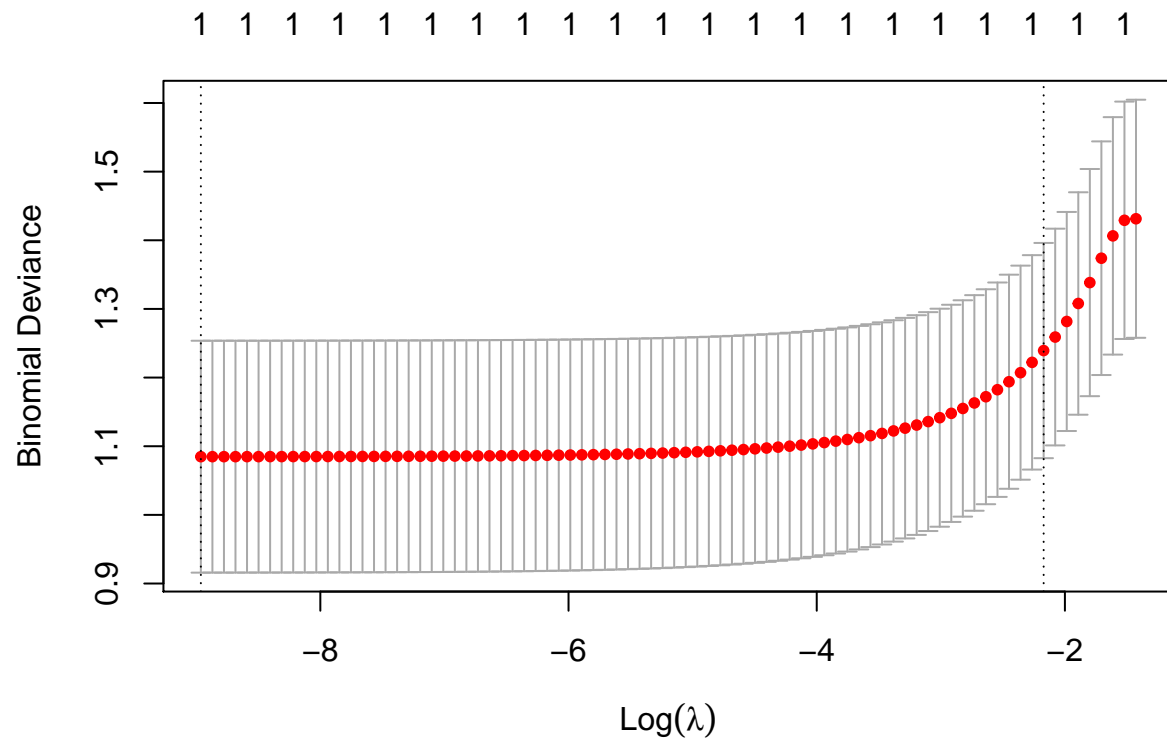
```
## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning in lognet(x, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning: Option grouped=FALSE enforced in cv.glmnet, since < 3 observations per
## fold
```

```
plot(cvfit1)
```



```
coef(cvfit, s = "lambda.min")
```

```
## 3 x 1 sparse Matrix of class "dgCMatrix"
##           1
## (Intercept) 0.9551772
## temp      .
## Diet      -2.4221092
```

```
coef(cvfit1, s = "lambda.min")
```

```
## 3 x 1 sparse Matrix of class "dgCMatrix"
##           1
## (Intercept) 2.738368
## temp      .
## Diet      -2.738975
```