Homework 3

Charles Hwang

2/17/2022

Charles Hwang

Dr. Whalen

STAT 410-001

17 February 2022

Problem 3.1

The link function provides the proper function to relate the response variable. The identity link is the link function $g(\mu) = \mu$ and it is seldom used with a binomial distribution because the range of the identity link, $[-\infty, \infty]$, is different than the range of the binomial parameter, which is defined as being 0 or 1.

Problem 3.2

Problem 3.2a

"For every 10 years since 1904, the model states the probability of a given starting pitcher pitching a complete game decreases by approximately 6.62 percent."

Problem 3.2b

$\hat{P}(Y = 1) = 0.6930 - 0.0662x = 0.6930 - 0.0662(12) = -0.1014$

The prediction from the logistic regression model is clearly more plausible. The predition from the linear model is outside the range of possible probabilities $[0, 1]$.

Problem 3.5

Problem 3.5a

```
rm(list=ls())
hc<-read.table("http://users.stat.ufl.edu/~aa/cat/data/Crabs.dat",header=TRUE)
ci<-lm(y~weight,data=hc)
summary(ci)
```

```
##
## Call:
## lm(formula = y ~ weight, data = hc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8878 -0.4683  0.1606  0.3704  0.6689
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.14487    0.14715  -0.984    0.326
## weight       0.32270    0.05876   5.492 1.42e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4447 on 171 degrees of freedom
## Multiple R-squared:  0.1499, Adjusted R-squared:  0.1449
## F-statistic: 30.16 on 1 and 171 DF,  p-value: 1.421e-07
```

Interpretation of *intercept*: "If the weight of a crab was 0 kilograms, the model states the probability of a crab having at least one satellite is approximately -0.1448709" (which obviously does not make sense).

Interpretation of *coefficient for weight*: "For every 1 kilogram increase in weight of a crab, the model states the probability of a crab having at least one satellite increases by approximately 0.3227033."

Problem 3.5b

```
as.numeric(ci$coefficients["(Intercept)"]+max(hc$weight)*ci$coefficients["weight"])
```

```
## [1] 1.533186
```

```
# We can see the model states the probability of a crab weighing 5.2 kilograms to have more
# than one satellite is approximately 1.533186 (which obviously does not make sense).
cl<-glm(y~weight,data=hc,family=binomial(link=logit))
summary(cl)
```

```
##
## Call:
## glm(formula = y ~ weight, family = binomial(link = logit), data = hc)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1108  -1.0749   0.5426   0.9122   1.6285
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.6947     0.8802  -4.198 2.70e-05 ***
## weight        1.8151     0.3767   4.819 1.45e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 225.76  on 172  degrees of freedom
## Residual deviance: 195.74  on 171  degrees of freedom
## AIC: 199.74
##
## Number of Fisher Scoring iterations: 4
```

```
eabx<-exp(cl$coefficients["(Intercept)"]+max(hc$weight)*cl$coefficients["weight"])
as.numeric(eabx/(1+eabx))
```

```
## [1] 0.9968084
```

Problem 3.6

Problem 3.6a

$P(Y=1)=\frac{e^{3.187-0.5901x}}{1+e^{3.187-0.5901x}}$

Interpretation of *direction*: "As one's political ideology gets more conservative, the probability of one being a Democrat decreases."

Problem 3.6b, 3.6c, 3.6d

```
library(car)
D<-c(5,18,19,25,7,7,2)
n<-D+c(1,3,1,11,10,11,1)
x<-1:7
p<-glm(D/n~x,family=binomial(link=logit),weights=n)
summary(p) # Problem 3.6b
```

```
##
## Call:
## glm(formula = D/n ~ x, family = binomial(link = logit), weights = n)
##
## Deviance Residuals:
##       1        2        3        4        5        6        7
## -0.8058  -0.3360   1.8917  -0.0154  -1.2160  -0.2041   1.3886
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.1870     0.7002   4.552 5.33e-06 ***
## x            -0.5901     0.1564  -3.772 0.000162 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 24.7983  on 6  degrees of freedom
## Residual deviance:  7.7894  on 5  degrees of freedom
## AIC: 30.516
##
## Number of Fisher Scoring iterations: 4
```

```
confint(p)["x",]
```

```
##      2.5 %     97.5 %
## -0.9158750 -0.2983235
```

```
cat("(",p$coefficients["x"]-qnorm(.975)*summary(p)$coefficients["x","Std. Error"],", ",p$coefficients["
```

```
## (-0.8967066, -0.2835034)
```

```
# This confidence interval is smaller than the displayed profile likelihood interval.
as.numeric(p$coefficients["x"]/summary(p)$coefficients["x","Std. Error"])
```

```
## [1] -3.772272
```

```
2*pnorm(as.numeric(p$coefficients["x"]/summary(p)$coefficients["x","Std. Error"]))
```

```
## [1] 0.0001617676
```

```r
# z = -3.772, p = 0.000162                                # Problem 3.6c
# We reject H0 at the alpha = 0.05 level. There is sufficient evidence (p = 0.000162) that
# the effect of x is significant in the model.
Anova(p)   # Problem 3.6d
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: D/n
##   LR Chisq Df Pr(>Chisq)
## x   17.009  1   3.72e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$\chi^2 = 17.0089141$, $p = 3.7204738 \times 10^{-5}$

We reject H_0 at the $\alpha = 0.05$ level. There is sufficient evidence (p = 0.0000372) that the effect of x is significant in the model.

Problem 3.6e

It took the model four (4) iterations to converge to the maximum likelihood value.

Problem 3.13

```r
cp<-glm(sat~weight,data=hc,family=poisson)
cp                                         # Problem 3.13a
```

```
##
## Call:  glm(formula = sat ~ weight, family = poisson, data = hc)
##
## Coefficients:
## (Intercept)       weight
##     -0.4284       0.5893
##
## Degrees of Freedom: 172 Total (i.e. Null);  171 Residual
## Null Deviance:       632.8
## Residual Deviance: 560.9     AIC: 920.2
```

```r
as.numeric(exp(cp$coefficients["(Intercept)"]+mean(hc$weight)*cp$coefficients["weight"]))
```

```
## [1] 2.73968
```

```r
as.numeric(exp(cp$coefficients["weight"])) # Problem 3.13b
```

```
## [1] 1.802734
```

```r
# The effect of weight on the model is quite large.
cat("(",cp$coefficients["weight"]-qnorm(.975)*summary(cp)$coefficients["weight","Std. Error"],", ",cp$co
```

```
## (0.4618742, 0.716734)
```

```r
cat("(",exp(cp$coefficients["weight"]-qnorm(.975)*summary(cp)$coefficients["weight","Std. Error"]),", "
```

```
## (1.587046, 2.047734)
```

95 percent confidence interval for β: (0.4618742, 0.716734)

95 percent confidence interval for the multiplicative effect of a 1-kg increase: (1.5870457, 2.0477345)