MATH 564 - Assignment6

2022-11-20

Mohammed Wasim RD(A20497053)

Problem 1 - Ex 14.9

```
data1<-read.table("http://www.cnachtsheim-text.csom.umn.edu/Kutner/Chapter%2014%20Data%20Sets/CH14PR09.t
colnames(data1)[1] ="Y"
colnames(data1)[2]="X"
head(data1)
```

```
##   Y   X
## 1 0 474
## 2 0 432
## 3 0 453
## 4 1 481
## 5 1 619
## 6 0 584
```

a)

```
lg = glm(Y ~ X, data = data1, family=binomial('logit'))
summary(lg)
```

```
##
## Call:
## glm(formula = Y ~ X, family = binomial("logit"), data = data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7845  -0.8350   0.5065   0.8371   1.7145
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -10.308925   4.376997  -2.355   0.0185 *
## X             0.018920   0.007877   2.402   0.0163 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 37.393  on 26  degrees of freedom
## Residual deviance: 29.242  on 25  degrees of freedom
## AIC: 33.242
##
## Number of Fisher Scoring iterations: 4
```

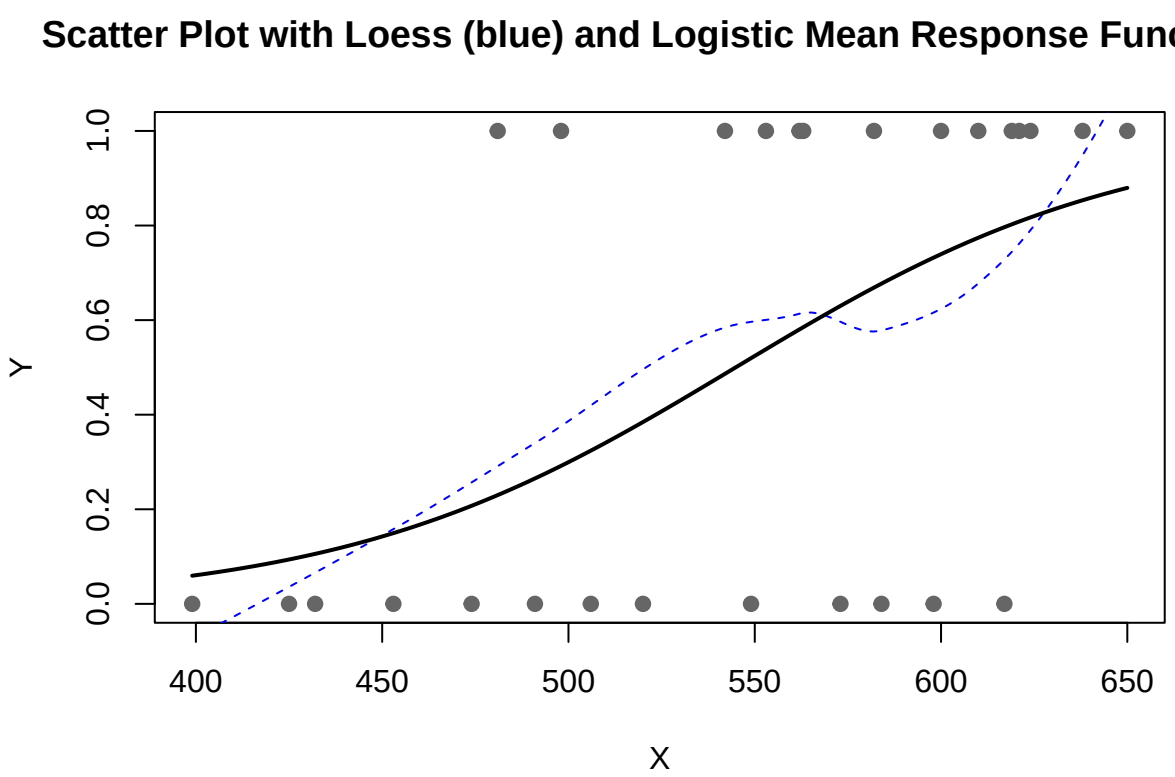
Fitted response function is

$b_0 = -10.308925$ $b_1 = 0.018920$ $\hat{\pi} = \frac{1}{1+e^{(10.308925-0.018920X)}}$

---

b)

```
xx <- with(data1, seq(min(X), max(X), len = 200))
plot(Y ~ X, data1, pch = 19, col = "gray40", xlab = "X", ylab = "Y")
lines(xx, predict(loess(Y ~ X, data1), data.frame(X = xx)), lty = 2, col = 'blue')
lines(xx, predict(lg, data.frame(X = xx), type = "resp"), lwd = 2)
title("Scatter Plot with Loess (blue) and Logistic Mean Response Functions")
```



The above plot seems to be good with low curves.

---

c)

```
expo_b1 = exp(0.018920)
cat("Exponent of beta 1 :", expo_b1)
```

```
## Exponent of beta 1 : 1.0191
```

The e^{β_1} *is* 1.0191, The odds ratio of able to perform in a task group(Y=1) versus unable to perform in a task group (Y=0) increase by 1.0191 times.

---

d)

```
cat("The estimated probability that employees with an emotional stability \n test scores of 550 will be
```

```
## The estimated probability that employees with an emotional stability
##  test scores of 550 will be able to perform in a task group is:  0.5242263
```

---

e)

```
xi= (log(.70/.30) - lg$coefficients[1])/lg$coefficients[2]
xi
```

```
## (Intercept)
##    589.6577
```

From the function : $\log \frac{\hat{\pi}}{1-\hat{\pi}} = \beta_0 + \beta_1 X$ knowing $\hat{\pi} = 0.70$, test score of $X = 589.65$ can be calculated to be. Therefore at least test score of 589.65 increase in will cause 70% of employee to able perform in the task group

---

Ex 14.10 (a)

```
probit_mean_response = glm(Y ~ X, data = data1, family=binomial('probit'))
summary(probit_mean_response)
```

```
##
## Call:
## glm(formula = Y ~ X, family = binomial("probit"), data = data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7940  -0.8336   0.4824   0.8380   1.7223
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.374398   2.464111  -2.587  0.00968 **
## X            0.011695   0.004437   2.636  0.00839 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 37.393  on 26  degrees of freedom
## Residual deviance: 29.102  on 25  degrees of freedom
## AIC: 33.102
##
## Number of Fisher Scoring iterations: 5
```

The probit link model function is : $\hat{\phi} = (-6.374398 + 0.011695X)$

If we look at the "probit" link model, it is better since it has a lower AIC and smaller Deviance.

Problem 2 - Ex 14.13

```
data2=read.table("http://www.cnachtsheim-text.csom.umn.edu/Kutner/Chapter%2014%20Data%20Sets/CH14PR13.t
colnames(data2)[1] ="Y"
colnames(data2)[2]="X1"
colnames(data2)[3] ="X2"
head(data2)
```

```
##   Y X1 X2
```

```
## 1 0 32  3
## 2 0 45  2
## 3 1 60  2
## 4 0 53  1
## 5 0 25  4
## 6 1 68  1
```

a)

```
car_logistic = glm(Y ~ X1 + X2, data=data2, family=binomial('logit'))
summary(car_logistic)
```

```
##
## Call:
## glm(formula = Y ~ X1 + X2, family = binomial("logit"), data = data2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6189  -0.8949  -0.5880   0.9653   2.0846
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.73931    2.10195  -2.255   0.0242 *
## X1           0.06773    0.02806   2.414   0.0158 *
## X2           0.59863    0.39007   1.535   0.1249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 44.987  on 32  degrees of freedom
## Residual deviance: 36.690  on 30  degrees of freedom
## AIC: 42.69
##
## Number of Fisher Scoring iterations: 4
```

$b_0 = -4.73931$ \$b_1 = 0.06773 \$ $b_2 = 0.59863$ $\hat{\pi} = \frac{1}{1+e^{(4.73931-0.06773X1-0.59863X2)}}$

---

b)

```
expo_b1 = exp(0.06773)
expo_b2 = exp(0.59863)
cat("Exponent of beta 1 :", expo_b1)
```

```
## Exponent of beta 1 : 1.070076
```

```
cat("\nExponent of beta 2 :", expo_b2)
```

```
##
## Exponent of beta 2 : 1.819624
```

e^{β_1} : 1.070076 e^{β_2} : 1.819624

Here beta 1 value says that for every every one unit increase in annual income, the odds ratio to family purchasing new car (Y=1) versus not purchasing (Y=0) increase by 1.070076 times

Here beta 2 value says that every one unit increase in current age of the oldest family automobile, the odds ratio to family purchasing new car (Y=1) versus not purchasing (Y=0) increase by 1.819624 times

---

c)

```
cat("The estimated probability that a family with annual income of \n$50 dollar and an oldest car of 3 y
```

```
## The estimated probability that a family with annual income of
## $50 dollar and an oldest car of 3 year will purchase a new car is:  0.6090245
```

Problem 3 - Ex 14.12

```
data3<-read.table("http://www.cnachtsheim-text.csom.umn.edu/Kutner/Chapter%2014%20Data%20Sets/CH14PR12.t
colnames(data3)[1] ="X"
colnames(data3)[2]="n"
colnames(data3)[3] ="Y"
head(data3)
```

```
##   X   n   Y
## 1 1 250  28
## 2 2 250  53
## 3 3 250  93
## 4 4 250 126
## 5 5 250 172
## 6 6 250 197
```

a)

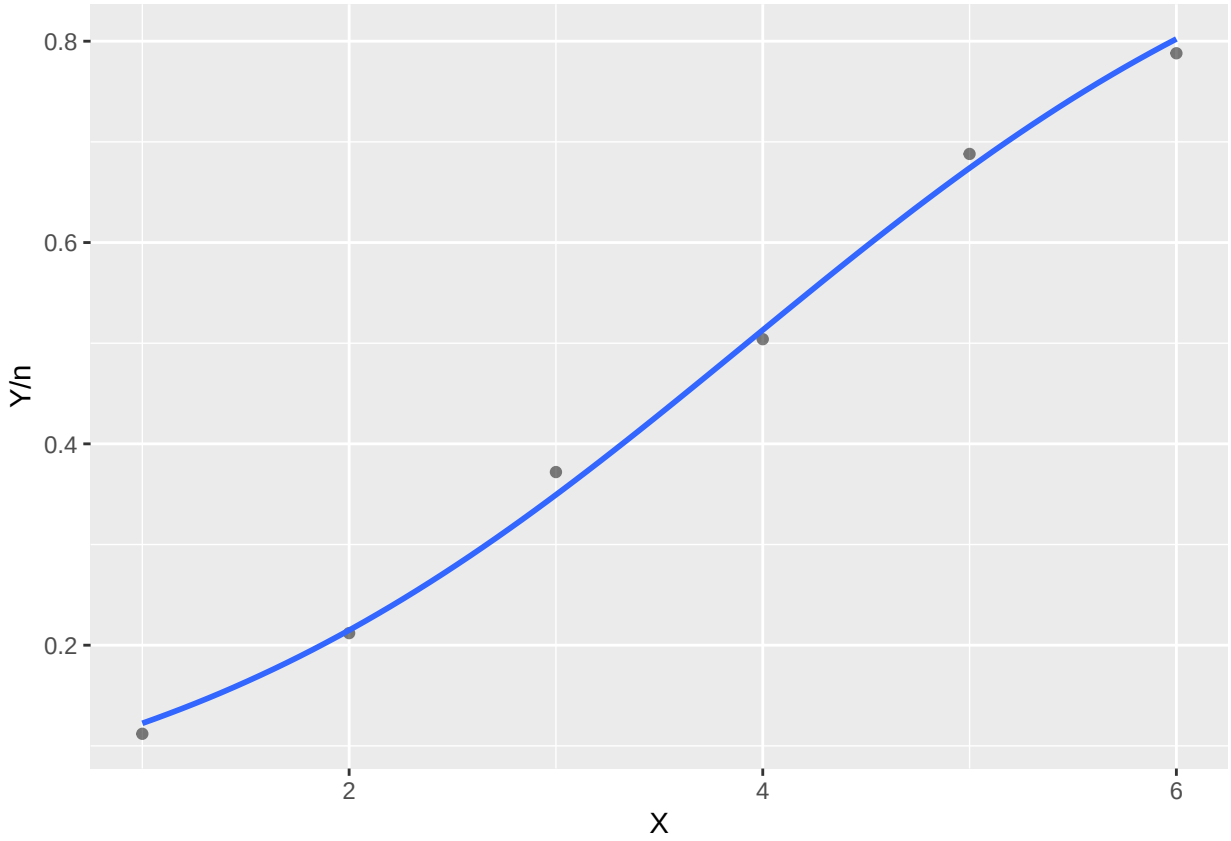
```
toxic_logistic = glm(Y/n ~ X, data=data3, family=quasibinomial(link = "logit"))
summary(toxic_logistic)
```

```
##
## Call:
## glm(formula = Y/n ~ X, family = quasibinomial(link = "logit"),
##     data = data3)
##
## Deviance Residuals:
##         1          2          3          4          5          6
## -0.032203  -0.007051   0.047185  -0.018146   0.030001  -0.035409
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.64367    0.09405  -28.11 9.53e-06 ***
## X            0.67399    0.02356   28.61 8.89e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.001451794)
##
##     Null deviance: 1.5322779  on 5  degrees of freedom
## Residual deviance: 0.0057964  on 4  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
library(ggplot2)
ggplot(data3, aes(x=X, y=Y/n)) + geom_point(alpha=.5) + stat_smooth(method="glm",se=FALSE, method.args=
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```



It seems like the logistic regression for this data seems appropriate. as we are having the curve.

---

b)

```
toxic_logistic = glm(Y/n ~ X, data=data3, family=quasibinomial(link = "logit"))
summary(toxic_logistic)
```

```
##
## Call:
## glm(formula = Y/n ~ X, family = quasibinomial(link = "logit"),
##     data = data3)
##
## Deviance Residuals:
##         1          2          3          4          5          6
## -0.032203  -0.007051   0.047185  -0.018146   0.030001  -0.035409
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.64367    0.09405  -28.11 9.53e-06 ***
## X            0.67399    0.02356   28.61 8.89e-06 ***
```

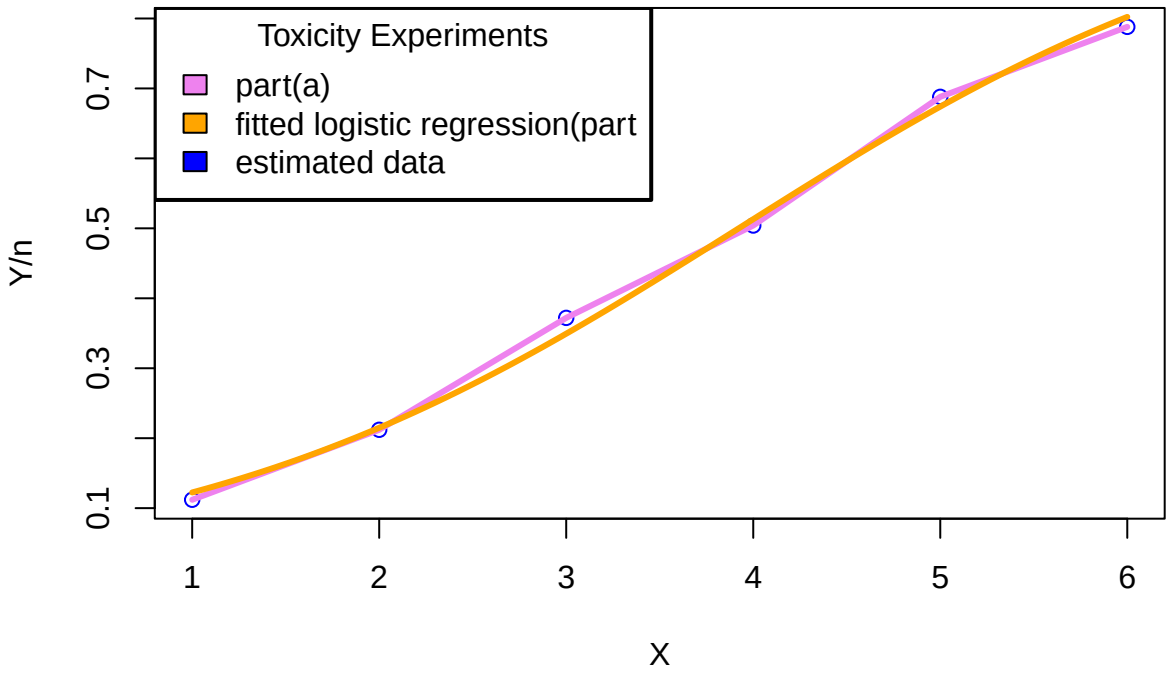
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.001451794)
##
##     Null deviance: 1.5322779  on 5  degrees of freedom
## Residual deviance: 0.0057964  on 4  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

$b_0 = -2.64367$ $b_1 = 0.67399$ $\hat{\pi} = \frac{1}{1+e^{(4.73931-0.06773X1)}}$

---

c)

```
toxic1 =  data.frame(X=seq(min(data3$X), max(data3$X),len=500))
toxic1$Y = predict(toxic_logistic,toxic1, type="response")
plot(Y/n ~ X, data = data3, col='blue')
lines(Y/n ~ X, data = data3, lwd=3, col='violet')
lines(Y ~ X, data=toxic1 , lwd= 3, col='orange')
legend(x = "topleft", box.col = "black", box.lwd = 2 , title="Toxicity Experiments", legend=c("part(a)"
```



Fitted Logistic Regression line (orange) seems to fit significantely good on the estimated datapoints.

d)

```
expo_b1 = exp(0.67399)
cat("Exponent of beta 1 :", expo_b1)
```

```
## Exponent of beta 1 : 1.96205
```

$e^{\beta_1} : e^{1.96205} = 1.96205$ is the increase in Odds of disease gain obtained increaseing in X by 1 unit

e)

```
cat("The estimated probability that an insect dies when the dose level is X = 3.5:", predict(toxic_logis
```

```
## The estimated probability that an insect dies when the dose level is X = 3.5: 0.4293018
```

---

f)

```
xi_toxic= (log(.50/.50) - toxic_logistic$coefficients[1])/toxic_logistic$coefficients[2]
xi_toxic
```

```
## (Intercept)
##    3.922409
```