STAT 3550 A2

Adrien Dinzey

29/03/2021

4.5

(a)

```
mydata <- read_excel("Ex4.5-Anand.xlsx")
mydata$P<-mydata$P/100
logitRegression <- glm(P ~ A + B + C + D, data = mydata,family = binomial)
summary(logitRegression)
```

```
##
## Call:
## glm(formula = P ~ A + B + C + D, family = binomial, data = mydata)
##
## Deviance Residuals:
##       Min         1Q     Median         3Q        Max
## -0.145350  -0.055786   0.005221   0.034967   0.120635
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.716e+01  3.818e+02   0.045    0.964
## A           -1.547e+01  3.399e+02  -0.046    0.964
## B           -3.850e-01  1.367e+00  -0.282    0.778
## C           -3.372e-03  6.798e-02  -0.050    0.960
## D           -9.227e-04  4.531e-02  -0.020    0.984
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 0.21144  on 26  degrees of freedom
## Residual deviance: 0.12661  on 22  degrees of freedom
## AIC: 18.179
##
## Number of Fisher Scoring iterations: 5
```

```
confint(logitRegression)
```

```
##                     2.5 %       97.5 %
## (Intercept) -769.01347150 810.96098285
## A           -722.41009423 684.25241752
```

```
## B             -3.30533495   2.35571597
## C             -0.14482805   0.13651265
## D             -0.09490356   0.09262787
```

(b)

```
mydata <- read_excel("Ex4.5-Anand.xlsx")
mydata$P<-mydata$P/100
logitRegression <- glm(P ~ A + B + C + D, data = mydata,family = binomial(link = "probit"))
summary(logitRegression)
```

```
##
## Call:
## glm(formula = P ~ A + B + C + D, family = binomial(link = "probit"),
##     data = mydata)
##
## Deviance Residuals:
##       Min         1Q     Median         3Q        Max
## -0.145570  -0.055366   0.004363   0.034427   0.118986
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  8.485e+00  2.067e+02   0.041    0.967
## A           -7.749e+00  1.840e+02  -0.042    0.966
## B           -2.082e-01  7.379e-01  -0.282    0.778
## C           -1.850e-03  3.680e-02  -0.050    0.960
## D           -4.672e-04  2.453e-02  -0.019    0.985
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 0.21144  on 26  degrees of freedom
## Residual deviance: 0.12674  on 22  degrees of freedom
## AIC: 18.179
##
## Number of Fisher Scoring iterations: 5
```

```
confint(logitRegression)
```

```
##                     2.5 %       97.5 %
## (Intercept) -399.08137124 417.60235944
## A           -372.00587496 355.14375104
## B             -1.69828648   1.23839458
## C             -0.07534342   0.07124907
## D             -0.04932177   0.04829991
```

(c)

we see that while they use the same number of Fisher Scoring iterations, the deviance is about the same and the z values for the estimates are also about the same, the actual estimates themselves are different. This is due to the fact that logit and probit have different link functions.

(d)

```
mydata <- read_excel("Ex4.5-Anand.xlsx")
mydata$P<-mydata$P/100
logitRegression <- glm(P ~ A + B + C + D, data = mydata,family = poisson)
summary(logitRegression)
```

```
##
## Call:
## glm(formula = P ~ A + B + C + D, family = poisson, data = mydata)
##
## Deviance Residuals:
##       Min         1Q     Median         3Q        Max
## -0.134560  -0.052425   0.005306   0.032427   0.111607
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  14.291213 353.413224   0.040    0.968
## A           -13.254240 314.651834  -0.042    0.966
## B            -0.330869   1.266998  -0.261    0.794
## C            -0.002889   0.062932  -0.046    0.963
## D            -0.000789   0.041948  -0.019    0.985
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 0.18037  on 26  degrees of freedom
## Residual deviance: 0.10746  on 22  degrees of freedom
## AIC: Inf
##
## Number of Fisher Scoring iterations: 5
```

```
confint(logitRegression)
```

```
##                    2.5 %       97.5 %
## (Intercept) -718.8919408 754.39432587
## A           -672.6717076 639.08233389
## B             -3.0624799   2.22372033
## C             -0.1348403   0.12751995
## D             -0.0884372   0.08643662
```

The poisson regression is very different from the previous two, , although it has a similar 5 number summary the estimates are extremely different. The deviance is somewhat different and the AIC is calculated as inf. Despite this, the z values are still very similar to the previous two techniques.

4.10

```
mydata <- read_excel("Ex4.10-popcorn.xlsx")
poissonReg <- glm(y ~ Temperature + Oil + Time, data = mydata,family = poisson)
summary(poissonReg)
```

```
##
## Call:
## glm(formula = y ~ Temperature + Oil + Time, family = poisson,
##     data = mydata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3497  -2.7886  -0.5767   1.3166   6.0752
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  8.243925   0.530997  15.525  < 2e-16 ***
## Temperature -0.346247   0.060457  -5.727 1.02e-08 ***
## Oil         -0.088511   0.059548  -1.486    0.137
## Time        -0.026742   0.004054  -6.597 4.19e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 213.45  on 14  degrees of freedom
## Residual deviance: 133.38  on 11  degrees of freedom
## AIC: 219.96
##
## Number of Fisher Scoring iterations: 5
```

```
confint(poissonReg)
```

```
## Waiting for profiling to be done...
```

```
##                   2.5 %      97.5 %
## (Intercept)  7.20278460  9.28465817
## Temperature -0.46525934 -0.22818908
## Oil         -0.20538067  0.02812055
## Time        -0.03472674 -0.01883148
```

4.18

(a)

```
mydata <- read_excel("4.18.xlsx")
logitRegression <- glm(y ~ x1+x2, data = mydata,family = binomial)
summary(logitRegression)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2, family = binomial, data = mydata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -1.5635  -0.8045  -0.1397   0.9535   1.7915
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.047e+00  4.674e+00  -1.508    0.132
## x1           7.382e-05  6.371e-05   1.159    0.247
## x2           9.879e-01  5.274e-01   1.873    0.061 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 27.726  on 19  degrees of freedom
## Residual deviance: 21.082  on 17  degrees of freedom
## AIC: 27.082
##
## Number of Fisher Scoring iterations: 5
```

```
confint(logitRegression)
```

```
## Waiting for profiling to be done...
```

```
##                     2.5 %       97.5 %
## (Intercept) -1.805544e+01 1.0275430082
## x1          -4.361540e-05 0.0002184223
## x2           1.544228e-01 2.2872127855
```

(b)

```
summaryLog <- summary(logitRegression)
1 - summaryLog$deviance / summaryLog$null.deviance
```

```
## [1] 0.239645
```

We see that R^2 is large enough that we fail to reject the null hypothesis and say this model is adequate based off its deviance.

(c)

```
exp(coef(logitRegression))
```

```
##  (Intercept)           x1           x2
## 0.0008699617 1.0000738195 2.6855513881
```

Since the odds ratio for x1 is close to 1, there is close to no change in the probability of success, while a change in x2 by 1 increases the probability of success by a factor of 2.68555

(d)

```
newData<-data.frame(x1=45000,x2=5)
predict(logitRegression,newData,type="response")
```

```
##         1
## 0.7710279
```

The probability of this is 0.7710279 according to this model.

(e)

```
mydata <- read_excel("4.18.xlsx")
logitRegression <- glm(y ~ x1*x2, data = mydata,family = binomial)
summary(logitRegression)
```

```
##
## Call:
## glm(formula = y ~ x1 * x2, family = binomial, data = mydata)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.63981  -0.62754  -0.05642   0.66213   1.85666
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.144e-01  6.394e+00   0.049    0.961
## x1          -1.411e-04  1.412e-04  -0.999    0.318
## x2          -2.462e+00  2.081e+00  -1.183    0.237
## x1:x2        1.014e-04  6.297e-05   1.610    0.107
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 27.726  on 19  degrees of freedom
## Residual deviance: 16.551  on 16  degrees of freedom
## AIC: 24.551
##
## Number of Fisher Scoring iterations: 6
```

```
exp(coef(logitRegression))
```

```
## (Intercept)          x1          x2       x1:x2
##  1.36936972  0.99985892  0.08529027  1.00010137
```

The interaction term has very little affect on the response variable due to the fact that it has an odds ratio very close to 1. Therefore, there is evidence this term is not required in the model.

(f)

The sigma values of x1 and x2 are 6.371e-05 and 5.274e-01 respectively.

(g)

```
mydata <- read_excel("4.18.xlsx")
logitRegression <- glm(y ~ x1+x2, data = mydata,family = binomial)
confint(logitRegression)
```

```
## Waiting for profiling to be done...
```

```
##                    2.5 %       97.5 %
## (Intercept) -1.805544e+01 1.0275430082
## x1          -4.361540e-05 0.0002184223
## x2           1.544228e-01 2.2872127855
```

4.19

(a)

```
data1 <-read_excel("4.19.xlsx")
poisReg<- glm( F~ Months, data = data1,family = poisson)
summary(poisReg)
```

```
##
## Call:
## glm(formula = F ~ Months, family = poisson, data = data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3106  -1.0114  -0.7003   0.4031   1.8813
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.71995    0.55770  -3.084  0.00204 **
## Months       0.13065    0.02433   5.370 7.88e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 44.167  on 14  degrees of freedom
## Residual deviance: 14.935  on 13  degrees of freedom
## AIC: 38.481
##
## Number of Fisher Scoring iterations: 5
```

(b)

```
summaryLog <- summary(poisReg)
1 - summaryLog$deviance / summaryLog$null.deviance
```

```
## [1] 0.6618563
```

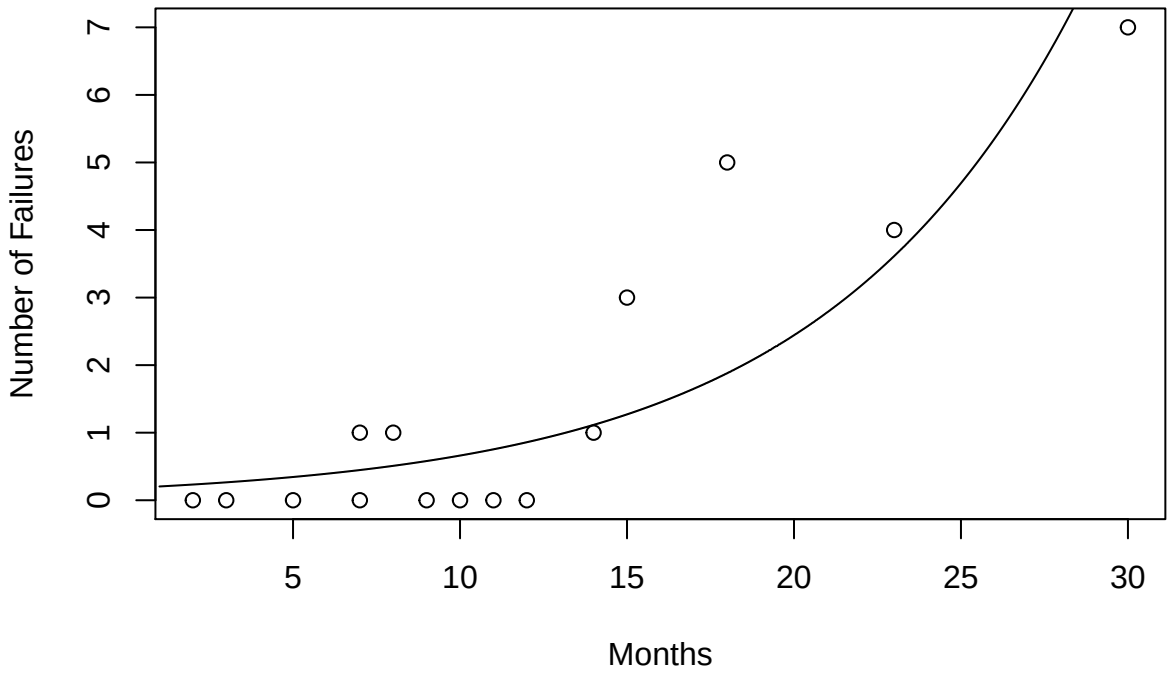
Because of this R^2 value, we fail to reject the null hypothesis at any reasonable confidence level, and so we can conclude that this model is a good fit.

(c)

```
range(data1$Months)
```

```
## [1]  2 30
```

```
xVals<-seq(1,35,0.001)
yVals<-predict(poisReg,list(Months=xVals),type="response")
plot(data1$Months,data1$F,xlab="Months",ylab="Number of Failures")
lines(xVals,yVals)
```



(d)

```
poisReg2<- glm( F~ Months+I(Months^2), data = data1,family = poisson)
summary(poisReg2)
```

```
##
## Call:
## glm(formula = F ~ Months + I(Months^2), family = poisson, data = data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3308  -0.8141  -0.3901   0.4821   1.2854
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.436107   1.705741  -2.601   0.0093 **
## Months       0.458657   0.179552   2.554   0.0106 *
## I(Months^2) -0.008259   0.004350  -1.899   0.0576 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 44.167  on 14  degrees of freedom
## Residual deviance: 10.769  on 12  degrees of freedom
## AIC: 36.315
##
## Number of Fisher Scoring iterations: 5
```

```
exp(coef(logitRegression))
```

```
##  (Intercept)           x1           x2
## 0.0008699617 1.0000738195 2.6855513881
```

Since the odds ratio for Months on its own is close to 1 it implies that it does not affect the probability of breaking very much, while the odds ratio of the quadratic term shows that it does have an effect on the probability of breaking meaning this is evidence that the term is required in the model.

(e)

```
regTermTest(poisReg,"Months", method=c("Wald"))
```

```
## Wald test for Months
##  in glm(formula = F ~ Months, family = poisson, data = data1)
## F =  28.83483  on  1  and  13  df: p= 0.00012758
```

The Wald Test F statistic is 28.83483 for Months.

(f)

```
confint(poisReg)
```

```
## Waiting for profiling to be done...
```

```
##                    2.5 %     97.5 %
## (Intercept) -2.92930837 -0.7217521
## Months       0.08388758  0.1803142
```