HW11

ASG

2022-11-22

```
setwd("/Users/animeshsengupta/Work Directory/DACSS/STAT625/Homeworks")
library(MASS)
library(alr4)  # loads the installed package into the workspace so you can use it
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: effects
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
library(summarytools)
library(ggplot2)
library(plotly)
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library(splines)
library(boot)
```

```
##
## Attaching package: 'boot'
```

```
## The following object is masked from 'package:car':
##
##     logit
```

```
library(sandwich)
library(plotly)
```

Answer 12.1

Answer 12.1.1

```
colnames(Blowdown)
```

```
## [1] "d"   "s"   "y"   "spp"
```

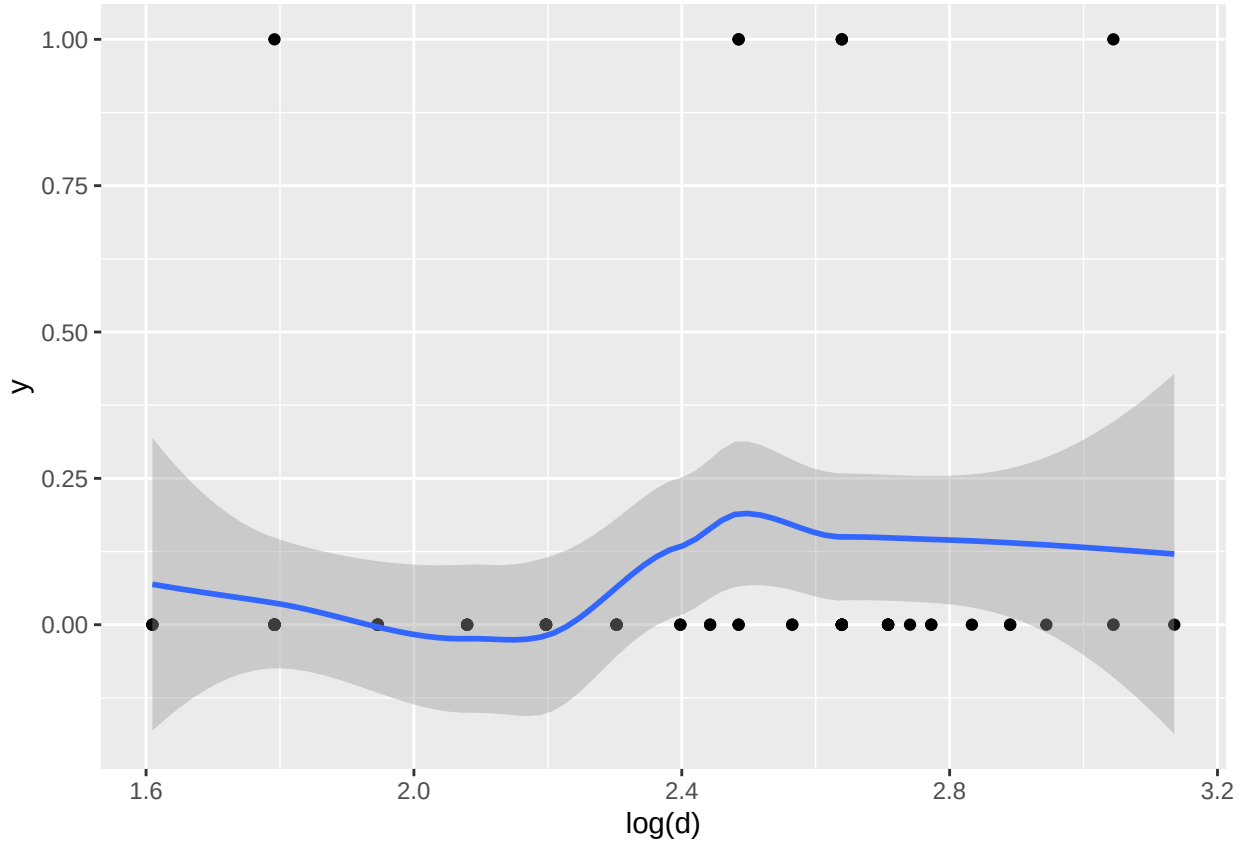
```
bd <- as.data.frame.matrix(xtabs(~ spp + y, Blowdown))
bd$total=bd$`0`+bd$`1`
bd$died=bd$`1`
bd
```

```
##                0   1 total died
## aspen        130 306   436  306
## balsam fir    69   6    75    6
## black spruce 426 233   659  233
## cedar        438 532   970  532
## jackpine     311  44   355   44
## paper birch   89 413   502  413
## red pine     407  90   497   90
## red maple    101  22   123   22
## black ash     11  38    49   38
```

12.1.2

```
bf<-Blowdown%>%filter(spp=="balsam fir")
ggplot(data=bf, aes(x=log(d), y=y)) +
        geom_point() +
        geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



As per the visualization , we can say that the response variable y only admits two values 0 and 1. Additionally plotting the smoother , we can clearly see that it shows a logistic curve. Hence fitting a logistic Generalised linear model would serve best.

12.1.3

```
bds <- glm(y ~ log(d), family=binomial,data=Blowdown%>%filter(spp=="black spruce"))
summary(bds)
```

```
##
## Call:
## glm(formula = y ~ log(d), family = binomial, data = Blowdown %>%
##     filter(spp == "black spruce"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5073  -0.7565  -0.4936   0.8096   2.3272
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -7.8925     0.6325  -12.48   <2e-16 ***
## log(d)        3.2643     0.2761   11.82   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 856.21  on 658  degrees of freedom
## Residual deviance: 655.24  on 657  degrees of freedom
## AIC: 659.24
##
## Number of Fisher Scoring iterations: 4
```

As seen from the summary, the Residual deviance and df are different from 49.891 and 33 df to 655.24 and 657 df.

12.1.4

```
bds2 <- glm(y ~ log(d)+I(log(d)^2), family=binomial,data=Blowdown%>%filter(spp=="black spruce"))
summary(bds2)
```

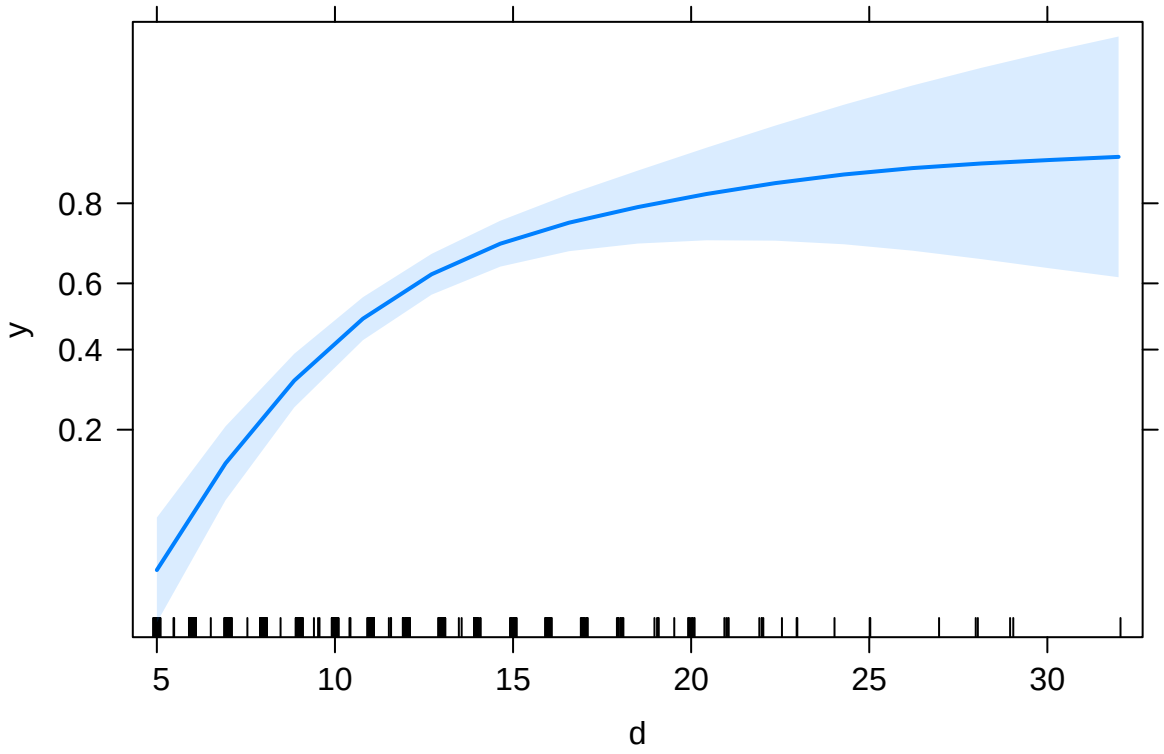
```
##
## Call:
## glm(formula = y ~ log(d) + I(log(d)^2), family = binomial, data = Blowdown %>%
##     filter(spp == "black spruce"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0193  -0.7841  -0.4515   0.8211   2.5095
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -14.3574     3.2948  -4.358 1.31e-05 ***
## log(d)        8.9726     2.8323   3.168  0.00154 **
## I(log(d)^2)  -1.2309     0.6004  -2.050  0.04034 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 856.21  on 658  degrees of freedom
## Residual deviance: 651.22  on 656  degrees of freedom
## AIC: 657.22
##
## Number of Fisher Scoring iterations: 5
```

```
anova(bds2,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: y
##
## Terms added sequentially (first to last)
##
```

```
##
##             Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                          658     856.21
## log(d)       1   200.97       657     655.24  < 2e-16 ***
## I(log(d)^2)  1     4.02       656     651.22  0.04496 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(Effect("d", bds2), main="")
```



Since z=0.0403 and g=0.044, z^2 and G^2 is definitely not same. the effect plots doesnt show a declining trend but declining probability cant be discounted, it can be plausible.

12.3

12.3.1

```
colnames(Downer)
```

```
## [1] "calving"  "daysrec"  "ck"       "ast"      "urea"     "pcv"      "inflamat"
## [8] "myopathy" "outcome"
```

```r
dw<-xtabs(~outcome+myopathy,data=Downer)
dw[2,]/colSums(dw)
```

```
##     absent    present
## 0.38582677 0.06315789
```

12.3.2

```r
dw1<-glm(outcome~myopathy,family=binomial,data=Downer)
summary(dw1)
```

```
##
## Call:
## glm(formula = outcome ~ myopathy, family = binomial, data = Downer)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9874  -0.9874  -0.3612  -0.3612   2.3504
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -0.4649     0.1823  -2.550   0.0108 *
## myopathypresent  -2.2320     0.4595  -4.858 1.19e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 248.57  on 221  degrees of freedom
## Residual deviance: 214.14  on 220  degrees of freedom
##   (213 observations deleted due to missingness)
## AIC: 218.14
##
## Number of Fisher Scoring iterations: 5
```

```r
confint(dw1)
```

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

```
##                       2.5 %     97.5 %
## (Intercept)     -0.8280172 -0.1115212
## myopathypresent -3.2323496 -1.4015854
```

```r
predict(dw1,data.frame(myopathy=factor(levels(Downer$myopathy))),type="response")
```

```
##          1          2
## 0.38582677 0.06315789
```

The intercept is when myopathy = 0, the estimated log-odds of survival . The coefficient for myopathy is the increase in log-odds when myopathy is present.

12.3.3

```
dw2 <- glm(outcome~log(ck),family=binomial, data=Downer)
summary(dw2)
```

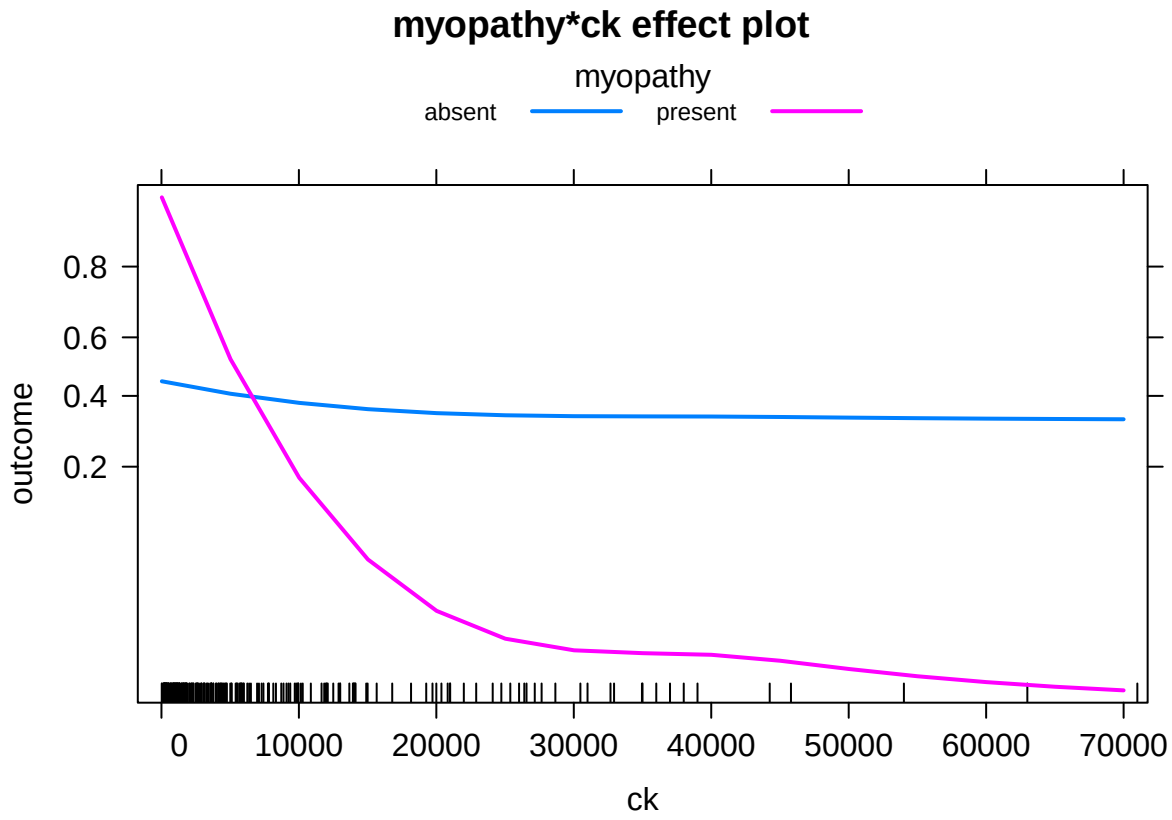
```
##
## Call:
## glm(formula = outcome ~ log(ck), family = binomial, data = Downer)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1337  -0.8811  -0.5608   1.0588   1.9935
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   4.0007     0.5809   6.887 5.69e-12 ***
## log(ck)      -0.6117     0.0793  -7.714 1.22e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 550.49  on 412  degrees of freedom
## Residual deviance: 475.18  on 411  degrees of freedom
##   (22 observations deleted due to missingness)
## AIC: 479.18
##
## Number of Fisher Scoring iterations: 3
```

12.3.4

```
dw3 <- glm(outcome~myopathy+log(ck)+myopathy:log(ck),family=binomial,data=Downer)
Anova(dw3)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: outcome
##                  LR Chisq Df Pr(>Chisq)
## myopathy          12.6153  1  0.0003826 ***
## log(ck)            1.3234  1  0.2499860
## myopathy:log(ck)   3.4204  1  0.0643943 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(Effect(c("myopathy", "ck"), dw3),multiline=TRUE)
```



12.9

12.9.1

```
colnames(AMSsurvey)
```

```
## [1] "type"    "sex"     "citizen" "count"   "count11"
```

```
AMS1 <- reshape(AMSsurvey, varying=c("count", "count11"), v.names="y",direction="long", times=c("08-09"
AMS1$type <- factor(AMS1$type, levels=levels(AMS1$type)[order(xtabs(y~type, AMS1))])
AMS1$year <- factor(AMS1$year)

ams1<-glm(y ~ (sex + citizen)*type, poisson, AMS1)
anova(ams1,test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: y
##
## Terms added sequentially (first to last)
```

```
##
##
##              Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                             47    1215.83
## sex           1   448.77        46     767.07 < 2.2e-16 ***
## citizen       1     7.32        45     759.74   0.00681 **
## type          5   533.76        40     225.98 < 2.2e-16 ***
## sex:type      5   108.10        35     117.88 < 2.2e-16 ***
## citizen:type  5    69.67        30      48.21 1.199e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#12.9.2

```
ams2 <- glm(y ~ (type + sex + citizen + year)^4, poisson, AMS1)
anova(ams2)
```

```
## Analysis of Deviance Table
##
## Model: poisson, link: log
##
## Response: y
##
## Terms added sequentially (first to last)
##
##
##                       Df Deviance Resid. Df Resid. Dev
## NULL                                     47    1215.83
## type                   5   533.76        42     682.07
## sex                    1   448.77        41     233.30
## citizen                1     7.32        40     225.98
## year                   1    14.91        39     211.07
## type:sex               5   108.10        34     102.97
## type:citizen           5    69.67        29      33.30
## type:year              5    11.04        24      22.26
## sex:citizen            1     2.58        23      19.68
## sex:year               1     2.51        22      17.17
## citizen:year           1     0.48        21      16.70
## type:sex:citizen       5     3.13        16      13.56
## type:sex:year          5     7.77        11       5.79
## type:citizen:year      5     4.24         6       1.55
## sex:citizen:year       1     0.17         5       1.39
## type:sex:citizen:year  5     1.39         0       0.00
```

```
ams3 <- update(ams2, ~ type*(sex + citizen + year))
Anova(ams3)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: y
##              LR Chisq Df Pr(>Chisq)
## type           533.76  5  < 2.2e-16 ***
## sex            448.77  1  < 2.2e-16 ***
```

```
## citizen          7.32  1  0.0068102 **
## year            14.91  1  0.0001128 ***
## type:sex       108.10  5  < 2.2e-16 ***
## type:citizen    69.67  5  1.199e-13 ***
## type:year       11.04  5  0.0505602 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

12.7.1

```
w1 <- glm(cbind(surv, m - surv) ~ class + age + sex,family=binomial, data=Whitestar)
summary(w1)
```

```
##
## Call:
## glm(formula = cbind(surv, m - surv) ~ class + age + sex, family = binomial,
##     data = Whitestar)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.1356  -1.7126   0.7812   2.6800   4.3833
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.1862     0.1586   7.480 7.40e-14 ***
## classfirst    0.8577     0.1573   5.451 5.00e-08 ***
## classsecond  -0.1604     0.1738  -0.923    0.356
## classthird   -0.9201     0.1486  -6.192 5.93e-10 ***
## agechild      1.0615     0.2440   4.350 1.36e-05 ***
## sexmale      -2.4201     0.1404 -17.236  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 671.96  on 13  degrees of freedom
## Residual deviance: 112.57  on  8  degrees of freedom
## AIC: 171.19
##
## Number of Fisher Scoring iterations: 5
```

Since all women survived apart from women in third class , we need to add an interaction term of sex*class.

12.7.2

```
w2 <- update(w1, ~(class + age + sex)^2)
Anova(w2)
```

```
## Analysis of Deviance Table (Type II tests)
```

```
##
## Response: cbind(surv, m - surv)
##           LR Chisq Df Pr(>Chisq)
## class       120.73  3  < 2.2e-16 ***
## age          20.34  1  6.486e-06 ***
## sex         359.37  1  < 2.2e-16 ***
## class:age    37.26  2  8.101e-09 ***
## class:sex    65.01  3  4.984e-14 ***
## age:sex       1.69  1     0.1942
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```