MATH 588

Final Exam

Md Ismail Hossain

5/03/2022

Question 1

a

```
library(Sleuth3)
library(readr)

fullBumpus <- read_table2("E:/NMT MS/Spring 22/MATH 588/Home_Work/Spring-2022---MATH-588-01-Advanced-Da

## 1 (a)

ttst1 = t.test(Weight~Survive,var.equal=TRUE,data = fullBumpus)
ttst1
```

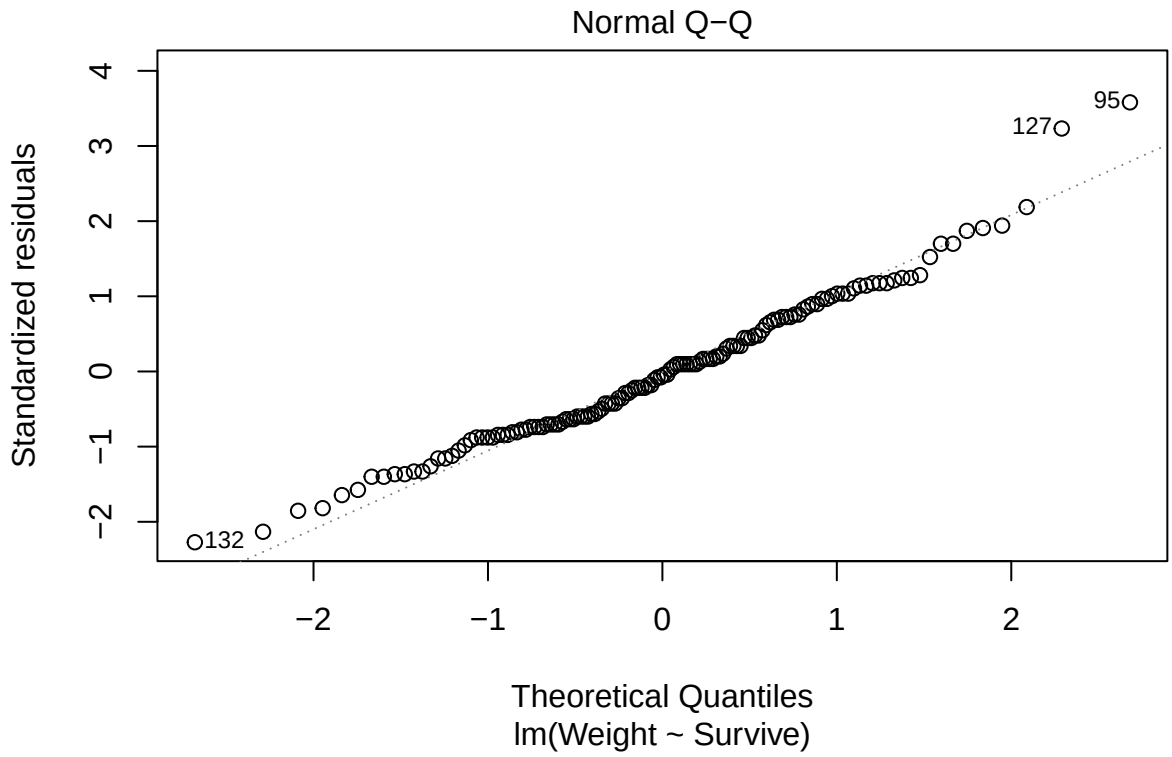
```
##
##  Two Sample t-test
##
## data:  Weight by Survive
## t = 2.6093, df = 134, p-value = 0.0101
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  0.1569291 1.1399459
## sample estimates:
## mean in group 0 mean in group 1
##        25.86094        25.21250
```

```
ttst2 = t.test(Weight~Survive,var.equal=FALSE,data = fullBumpus)
ttst2
```

```
##
##  Welch Two Sample t-test
##
## data:  Weight by Survive
## t = 2.5703, df = 117.95, p-value = 0.01141
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  0.1488463 1.1480287
## sample estimates:
## mean in group 0 mean in group 1
##        25.86094        25.21250
```

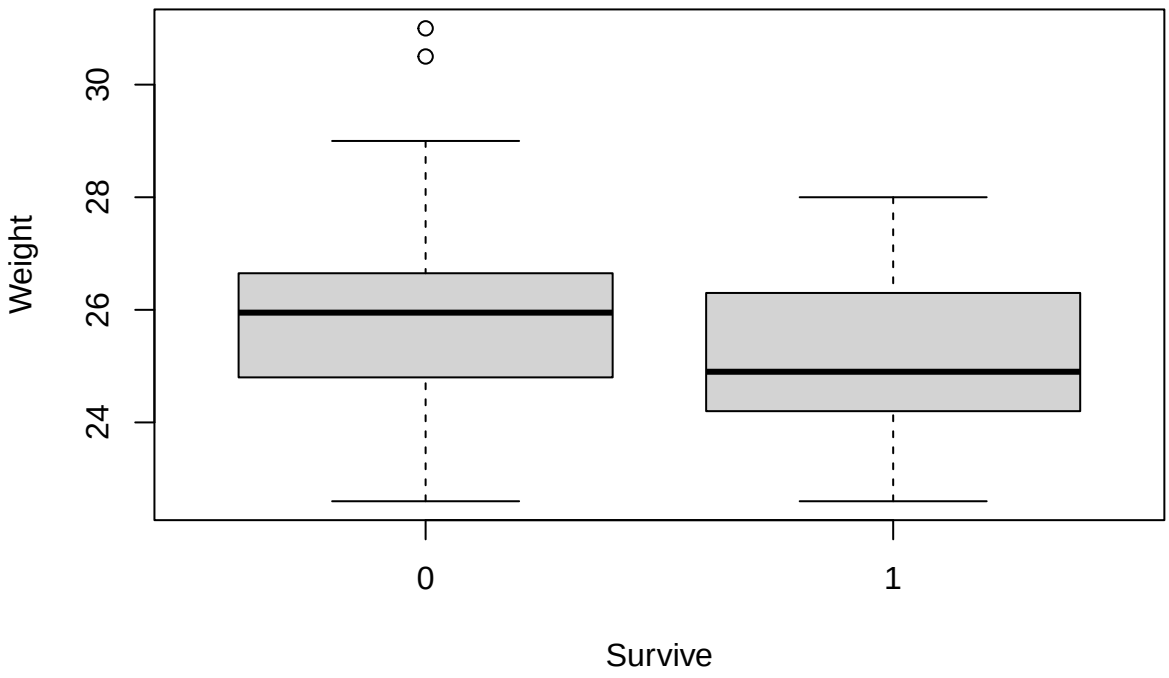
b

```
fit <- lm(Weight~Survive,data = fullBumpus)
plot(fit,2)
```



c

```
boxplot(Weight~Survive,data=fullBumpus)
```




```
#Showing IQR
aggregate(Weight~Survive,fullBumpus,IQR)
```

```
##   Survive Weight
## 1       0  1.775
## 2       1  2.100
```

Question 3

a

```
head(ex2220)
```

```
##   Exposure YearsAfter AtRisk Deaths
## 1        0       0to7    262     10
## 2        0      8to11    243     12
## 3        0     12to15    240     19
## 4        0     16to19    237     31
## 5        0     20to23    233     35
## 6        0     24to27    227     48
```

```
ex2220$Exposer_sq <- ex2220$Exposure^2
m1 <- glm(Deaths ~ log(AtRisk)+factor(YearsAfter)+Exposure+Exposer_sq , family="poisson", data=ex2220)
summary(m1)
```

```
##
## Call:
```

```
## glm(formula = Deaths ~ log(AtRisk) + factor(YearsAfter) + Exposure +
##     Exposer_sq, family = "poisson", data = ex2220)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.35893  -0.83139  -0.09575   0.47651   2.68348
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -3.299e+00  6.464e-01  -5.104 3.32e-07 ***
## log(AtRisk)                1.006e+00  1.049e-01   9.593  < 2e-16 ***
## factor(YearsAfter)12to15   5.526e-01  2.373e-01   2.329   0.0199 *
## factor(YearsAfter)16to19   1.249e+00  2.135e-01   5.851 4.90e-09 ***
## factor(YearsAfter)20to23   1.405e+00  2.104e-01   6.677 2.44e-11 ***
## factor(YearsAfter)24to27   1.738e+00  2.044e-01   8.503  < 2e-16 ***
## factor(YearsAfter)28to31   2.033e+00  2.006e-01  10.135  < 2e-16 ***
## factor(YearsAfter)8to11    2.337e-01  2.529e-01   0.924   0.3555
## Exposure                   4.573e-03  2.702e-03   1.693   0.0905 .
## Exposer_sq                -7.638e-06  5.368e-06  -1.423   0.1548
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 894.448  on 41  degrees of freedom
## Residual deviance:  46.686  on 32  degrees of freedom
## AIC: 216.44
##
## Number of Fisher Scoring iterations: 5
```

b

```
head(ex2220,20)
```

```
##    Exposure YearsAfter AtRisk Deaths Exposer_sq
## 1         0       0to7    262     10          0
## 2         0      8to11    243     12          0
## 3         0     12to15    240     19          0
## 4         0     16to19    237     31          0
## 5         0     20to23    233     35          0
## 6         0     24to27    227     48          0
## 7         0     28to31    220     73          0
## 8        25       0to7    313     17        625
## 9        25      8to11    290     17        625
## 10       25     12to15    285     17        625
## 11       25     16to19    280     47        625
## 12       25     20to23    275     50        625
## 13       25     24to27    269     65        625
## 14       25     28to31    262     71        625
## 15       75       0to7     38      0       5625
## 16       75      8to11     36      2       5625
## 17       75     12to15     35      1       5625
## 18       75     16to19     34      5       5625
## 19       75     20to23     34      8       5625
```

```
## 20       75     24to27     33      7       5625
```

```
ex2220$Time <- ifelse(ex2220$YearsAfter ==  "0to7",3.5,
                      ifelse(ex2220$YearsAfter == "8to11",9.5,
                             ifelse(ex2220$YearsAfter == "12to15",13.5,
                                    ifelse(ex2220$YearsAfter == "16to19",17.5,
                                           ifelse(ex2220$YearsAfter == "20to23",21.5,
                                                  ifelse(ex2220$YearsAfter == "24to27",25.5,
                                                         ifelse(ex2220$YearsAfter == "28to31",29.5,NA
                                                                ))))))))

m2 <- glm(Deaths ~ log(AtRisk)+log(Time)+Exposure+Exposer_sq , family="poisson", data=ex2220)

summary(m2)
```

```
##
## Call:
## glm(formula = Deaths ~ log(AtRisk) + log(Time) + Exposure + Exposer_sq,
##     family = "poisson", data = ex2220)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7161  -0.8760  -0.3090   0.4216   3.4267
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.501e+00  6.946e-01  -7.919 2.38e-15 ***
## log(AtRisk)  9.977e-01  1.048e-01   9.520  < 2e-16 ***
## log(Time)    1.223e+00  9.469e-02  12.917  < 2e-16 ***
## Exposure     4.387e-03  2.700e-03   1.625    0.104
## Exposer_sq  -7.338e-06  5.366e-06  -1.367    0.171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 894.448  on 41  degrees of freedom
## Residual deviance:  73.679  on 37  degrees of freedom
## AIC: 233.43
##
## Number of Fisher Scoring iterations: 5
```

c

```
m3 <- glm(Deaths ~ log(AtRisk)+log(Time)+log(Time)*Exposure+Exposure+Exposer_sq , family="poisson", data
summary(m3)
```

```
##
## Call:
## glm(formula = Deaths ~ log(AtRisk) + log(Time) + log(Time) *
##     Exposure + Exposure + Exposer_sq, family = "poisson", data = ex2220)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7705  -0.8517  -0.1960   0.5409   3.2862
```

```
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -5.328e+00  7.097e-01  -7.508 6.01e-14 ***
## log(AtRisk)         9.983e-01  1.048e-01   9.526  < 2e-16 ***
## log(Time)           1.164e+00  1.074e-01  10.840  < 2e-16 ***
## Exposure            4.382e-04  4.552e-03   0.096    0.923
## Exposer_sq         -7.520e-06  5.375e-06  -1.399    0.162
## log(Time):Exposure  1.324e-03  1.232e-03   1.075    0.282
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 894.448  on 41  degrees of freedom
## Residual deviance:  72.426  on 36  degrees of freedom
## AIC: 234.17
##
## Number of Fisher Scoring iterations: 5
```

d

```
with(m1, cbind(res.deviance = deviance, df = df.residual,
  p = pchisq(deviance, df.residual, lower.tail=FALSE)))
```

```
##      res.deviance df          p
## [1,]     46.68643 32 0.04523257
```

```
with(m2, cbind(res.deviance = deviance, df = df.residual,
  p = pchisq(deviance, df.residual, lower.tail=FALSE)))
```

```
##      res.deviance df            p
## [1,]     73.67922 37 0.0003150971
```

```
with(m3, cbind(res.deviance = deviance, df = df.residual,
  p = pchisq(deviance, df.residual, lower.tail=FALSE)))
```

```
##      res.deviance df            p
## [1,]     72.42569 36 0.0003027446
```

```
require(MASS)
exp(confint(m1))
```

```
##                                2.5 %     97.5 %
## (Intercept)              0.01027875  0.1296804
## log(AtRisk)              2.22880558  3.3624848
## factor(YearsAfter)12to15 1.09806011  2.7955763
## factor(YearsAfter)16to19 2.32608171  5.3898883
## factor(YearsAfter)20to23 2.73712720  6.2651236
## factor(YearsAfter)24to27 3.86982994  8.6486277
## factor(YearsAfter)28to31 5.24299933 11.5441459
## factor(YearsAfter)8to11  0.77007982  2.0858589
## Exposure                 0.99928887  1.0099350
## Exposer_sq               0.99998180  1.0000029
```

Question 4

a

```
Dose <- c(0.0794,0.1000,0.1259,0.1413,0.1500,0.1588,0.1778,0.1995,0.2239,0.2512,0.2818,0.3162)
Group <- c(1:12)
Died <- c(1,2,1,0,1,2,4,6,4,5,5,8)
Survived <- c(9,8,9,10,9,8,6,4,6,5,5,2)
Total <- rep(10,12)

dat4 <- data.frame(Group,Dose,Died,Survived,Total)
dat4$Proportion <- dat4$Died/dat4$Total

head(dat4)
```

```
##   Group   Dose Died Survived Total Proportion
## 1     1 0.0794    1        9    10        0.1
## 2     2 0.1000    2        8    10        0.2
## 3     3 0.1259    1        9    10        0.1
## 4     4 0.1413    0       10    10        0.0
## 5     5 0.1500    1        9    10        0.1
## 6     6 0.1588    2        8    10        0.2
```

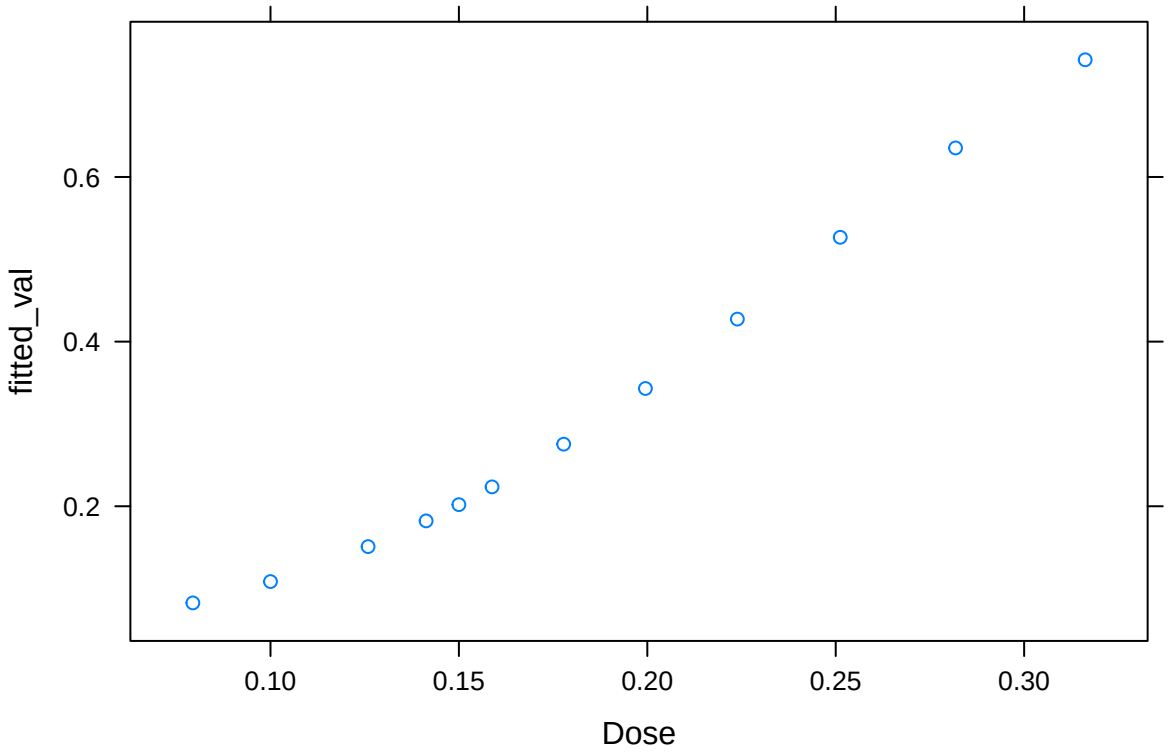
```
attach(dat4)
y <- cbind(Died,Survived)
fit1 <- glm (y ~ Dose, data = dat4, family = binomial)
summary (fit1)
```

```
##
## Call:
## glm(formula = y ~ Dose, family = binomial, data = dat4)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0057  -0.5740  -0.1724   0.5322   1.6550
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.5697     0.7053  -5.061 4.17e-07 ***
## Dose         14.6369     3.3325   4.392 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 33.704  on 11  degrees of freedom
## Residual deviance: 10.254  on 10  degrees of freedom
## AIC: 41.034
##
## Number of Fisher Scoring iterations: 4
```

b

```
dat4$fitted_val <- fitted(fit1)
```

```
require(lattice)
xyplot (fitted_val~Dose, data = dat4)
```



d

```
# Goodness of fit test:
df       <- 10
deviance <- 10.254
p_val    <- pchisq(deviance, df=df, lower.tail=FALSE)
p_val
```

```
## [1] 0.4184988
```