Booklet of Code and Output
for
STAD29/STA 1007 Midterm Exam

List of Figures in this document by page:

List of Figures

```r
nba=read.table("nba-attendance.txt",header=T)
nba

##              team wins market.share finals attendance
## 1         Atlanta   47        2.070     no      16478
## 2          Boston   62        2.105    yes      18624
## 3       Charlotte   35        0.981     no      14526
## 4         Chicago   41        3.052     no      21197
## 5       Cleveland   66        1.332    yes      20010
## 6          Dallas   50        2.175    yes      20042
## 7          Denver   54        1.332     no      17223
## 8         Detroit   39        1.684    yes      21877
## 9     GoldenState   29        2.164     no      18942
## 10        Houston   53        1.840     no      17482
## 11        Indiana   36        0.974     no      14182
## 12    LAClippers   19        4.940     no      16170
## 13       LALakers   65        4.940    yes      18997
## 14        Memphis   24        0.589     no      12745
## 15          Miami   43        1.352    yes      18229
## 16      Milwaukee   34        0.791     no      15389
## 17      Minnesota   24        1.512     no      14505
## 18      NewJersey   34        6.495     no      15147
## 19     NewOrleans   49        0.527     no      16968
## 20        NewYork   32        6.495     no      19287
## 21   OklahomaCity   23        0.600     no      18693
## 22        Orlando   59        1.281    yes      17043
## 23   Philadelphia   41        2.578     no      15802
## 24        Phoenix   46        1.622     no      18422
## 25       Portland   54        1.027     no      20524
## 26     Sacramento   17        1.223     no      12571
## 27     SanAntonio   54        0.715    yes      18269
## 28        Toronto   33        4.900     no      18773
## 29           Utah   48        0.803     no      19903
## 30     Washington   19        2.028     no      16612
```

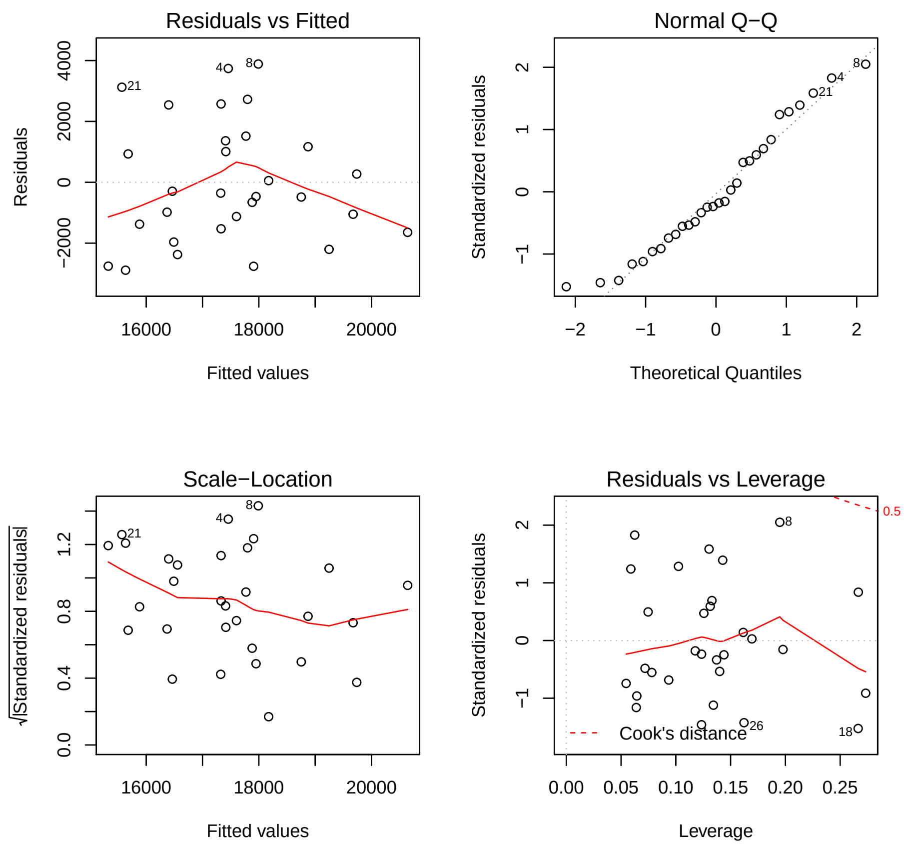
Figure 1: NBA attendance data

```r
nba.1=lm(attendance~wins+market.share+finals,data=nba)
summary(nba.1)

##
## Call:
## lm(formula = attendance ~ wins + market.share + finals, data = nba)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2888.4 -1490.0  -411.6  1315.2  3886.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13836.50    1473.50   9.390 7.73e-10 ***
## wins            68.26      34.87   1.958   0.0611 .
## market.share   269.36     232.72   1.157   0.2576
## finalsyes     1037.88    1088.32   0.954   0.3490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2113 on 26 degrees of freedom
## Multiple R-squared:  0.3049,Adjusted R-squared:  0.2247
## F-statistic: 3.802 on 3 and 26 DF,  p-value: 0.02199
```

Figure 2: Regression model for NBA attendances

```r
par(mfrow=c(2,2))
plot(nba.1)
```




```r
par(mfrow=c(1,1))
```

Figure 3: Residual plots for NBA attendances

```r
ozone=read.table("inhibit.txt",header=T)
ozone

##    inhibit.deep uvb.deep inhibit.surface uvb.surface
## 1           0.0     0.00             7.0        0.01
## 2           1.0     0.00             7.0        0.02
## 3           6.0     0.01             7.0        0.03
## 4           9.5     0.01             9.0        0.04
## 5          10.0     0.00            11.0        0.03
## 6          14.0     0.01            12.5        0.03
## 7          20.0     0.03            21.0        0.04
## 8          25.0     0.02              NA          NA
## 9          39.0     0.03              NA          NA
## 10         59.0     0.03              NA          NA

suppressMessages(library(dplyr))

## Warning: package 'dplyr' was built under R version 3.5.2
```

Figure 4: Ozone layer data

```r
ozone %>% mutate(inhibit=inhibit.deep,uvb=uvb.deep) %>%
  select(c(inhibit,uvb)) -> deep
ozone %>% mutate(inhibit=inhibit.surface,uvb=uvb.surface) %>%
  select(c(inhibit,uvb)) -> surface
```

Figure 5: Ozone layer data processing part 1

```
v=bind_rows(deep=deep,surface=surface,.id="depth")
v

##      depth inhibit  uvb
## 1     deep     0.0 0.00
## 2     deep     1.0 0.00
## 3     deep     6.0 0.01
## 4     deep     9.5 0.01
## 5     deep    10.0 0.00
## 6     deep    14.0 0.01
## 7     deep    20.0 0.03
## 8     deep    25.0 0.02
## 9     deep    39.0 0.03
## 10    deep    59.0 0.03
## 11 surface     7.0 0.01
## 12 surface     7.0 0.02
## 13 surface     7.0 0.03
## 14 surface     9.0 0.04
## 15 surface    11.0 0.03
## 16 surface    12.5 0.03
## 17 surface    21.0 0.04
## 18 surface      NA   NA
## 19 surface      NA   NA
## 20 surface      NA   NA
```

Figure 6: Ozone layer data processing part 2

```r
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.3

ggplot(v,aes(x=uvb,y=inhibit,colour=depth))+
  geom_point()+geom_smooth(method="lm")

## Warning: Removed 3 rows containing non-finite values
(stat_smooth).
## Warning: Removed 3 rows containing missing values
(geom_point).
```

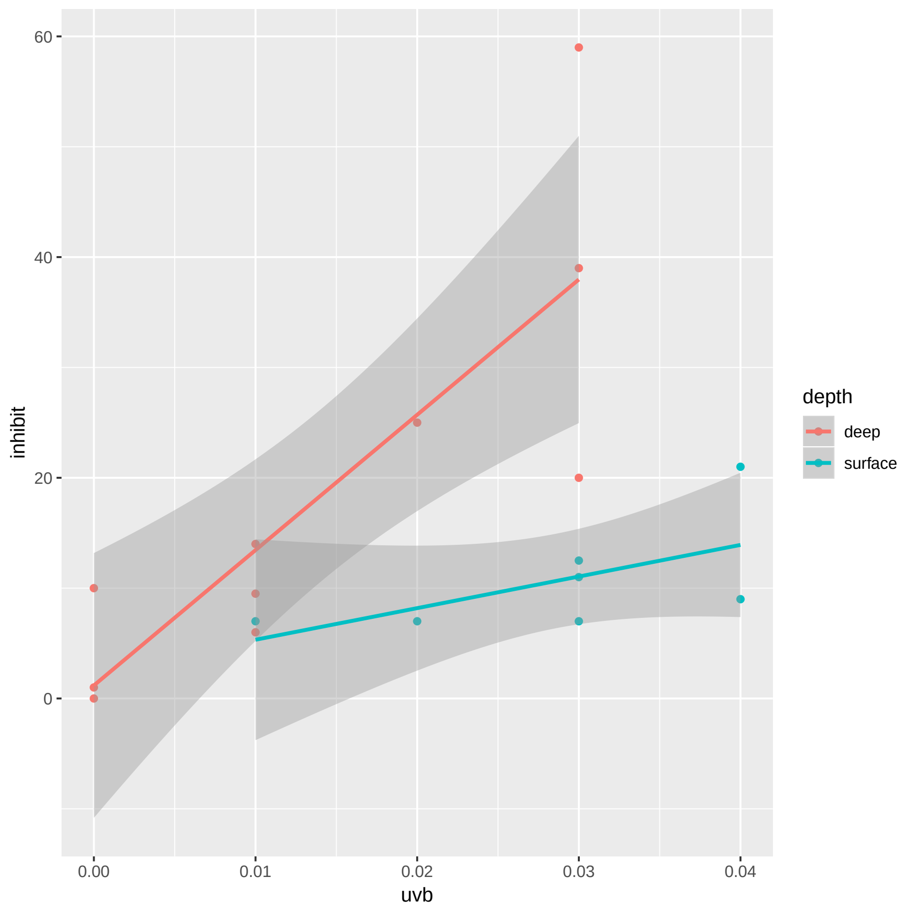


Figure 7: Ozone layer scatterplot

```r
nzbirds=read.table("nzbird.txt",header=T)
str(nzbirds)
```

```
## 'data.frame': 67 obs. of  15 variables:
##  $ Species: Factor w/ 67 levels "Aca_cann","Aca_flam",..: 29 28 23 14 13 18 19 12 10 11 .
##  $ Status : int  1 1 1 0 0 1 0 1 0 0 ...
##  $ Length : int  1520 1250 870 720 820 770 50 570 580 480 ...
##  $ Mass   : num  9600 5000 3360 2517 3170 ...
##  $ Range  : num  1.21 0.56 0.07 1.1 3.45 2.96 0.01 9.01 7.9 4.33 ...
##  $ Migr   : int  1 1 1 3 3 2 1 2 3 3 ...
##  $ Insect : int  12 0 0 12 0 0 0 6 6 0 ...
##  $ Diet   : int  2 1 1 2 1 1 1 2 2 1 ...
##  $ Clutch : num  6 6 4 3.8 5.9 5.9 4 12.6 8.3 8.7 ...
##  $ Broods : int  1 1 1 1 1 1 2 1 1 1 ...
##  $ Wood   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Upland : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Water  : int  1 1 1 1 1 1 0 1 1 1 ...
##  $ Release: int  6 10 3 1 2 10 1 17 3 5 ...
##  $ Indiv  : int  29 85 8 10 7 60 2 1539 102 32 ...
```

Figure 8: New Zealand bird data

```
nzbirds.1=glm(Status~Length+Mass+Range+Migr+Insect+Diet+
  Clutch+Broods+Wood+Upland+Water+Release+Indiv,
  data=nzbirds,family="binomial")
summary(nzbirds.1)

##
## Call:
## glm(formula = Status ~ Length + Mass + Range + Migr + Insect +
##     Diet + Clutch + Broods + Wood + Upland + Water + Release +
##     Indiv, family = "binomial", data = nzbirds)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.56830  -0.25666  -0.04783   0.10892   2.72372
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.338010   5.716762  -1.109   0.2676
## Length      -0.002815   0.005317  -0.529   0.5965
## Mass         0.002668   0.001674   1.594   0.1110
## Range       -0.131607   0.350234  -0.376   0.7071
## Migr        -2.043545   1.115824  -1.831   0.0670 .
## Insect       0.147992   0.212364   0.697   0.4859
## Diet         2.028505   1.883201   1.077   0.2814
## Clutch       0.079380   0.268305   0.296   0.7673
## Broods       0.021770   0.928327   0.023   0.9813
## Wood         2.490210   1.641601   1.517   0.1293
## Upland      -4.713474   2.864827  -1.645   0.0999 .
## Water        0.234944   2.670193   0.088   0.9299
## Release     -0.012916   0.193211  -0.067   0.9467
## Indiv        0.015926   0.008324   1.913   0.0557 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 90.343  on 66  degrees of freedom
## Residual deviance: 26.496  on 53  degrees of freedom
## AIC: 54.496
##
## Number of Fisher Scoring iterations: 8
```

Figure 9: Analysis 1 of NZ bird data

```r
nzbirds.2=update(nzbirds.1,.~.-Length-Range-Insect-Clutch
  -Broods-Water-Release)
summary(nzbirds.2)

##
## Call:
## glm(formula = Status ~ Mass + Migr + Diet + Wood + Upland + Indiv,
##     family = "binomial", data = nzbirds)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.62601  -0.32089  -0.07121   0.10296   2.73035
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.1367680  2.7210950  -1.888 0.059058 .
## Mass         0.0019131  0.0007512   2.547 0.010876 *
## Migr        -1.9596476  0.9658570  -2.029 0.042466 *
## Diet         1.8677300  0.9987152   1.870 0.061465 .
## Wood         2.3002854  1.3579902   1.694 0.090286 .
## Upland      -5.1221630  2.4552556  -2.086 0.036960 *
## Indiv        0.0157164  0.0046631   3.370 0.000751 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 90.343  on 66  degrees of freedom
## Residual deviance: 28.052  on 60  degrees of freedom
## AIC: 42.052
##
## Number of Fisher Scoring iterations: 7
```

Figure 10: Analysis 2 of NZ bird data

```r
anova(nzbirds.2,nzbirds.1,test="Chisq")

## Analysis of Deviance Table
##
## Model 1: Status ~ Mass + Migr + Diet + Wood + Upland + Indiv
## Model 2: Status ~ Length + Mass + Range + Migr + Insect + Diet + Clutch +
##     Broods + Wood + Upland + Water + Release + Indiv
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        60     28.052
## 2        53     26.496  7   1.5554   0.9803
```

Figure 11: Analysis 3 of NZ bird data

```r
space.shuttle=read.table("space-shuttle.txt",header=T)
str(space.shuttle)

## 'data.frame': 23 obs. of  5 variables:
##  $ flight     : int  1 2 3 5 6 7 8 9 10 11 ...
##  $ distress   : Factor w/ 3 levels "0","1-
2","3+": 1 2 1 1 2 1 1 1 2 3 ...
##  $ temp       : int  66 70 69 68 67 72 73 70 57 63 ...
##  $ date       : int  7772 7986 8116 8350 8494 8569 8642 8732 8799 8862 ...
##  $ z.computed.: num  14.1 14.1 14.7 15.6 16.3 ...
```

Figure 12: Space shuttle data (summary)

```r
library(MASS)

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

space.shuttle.1=polr(distress~temp+date,data=space.shuttle)
space.shuttle.2=polr(distress~date,data=space.shuttle)
space.shuttle.3=polr(distress~temp,data=space.shuttle)
```

Figure 13: Space shuttle part 1

```r
anova(space.shuttle.2,space.shuttle.1)

## Likelihood ratio tests of ordinal regression models
##
## Response: distress
##         Model Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
## 1        date        20   43.71774
## 2 temp + date        19   37.59413 1 vs 2     1 6.123617 0.01333876
```

Figure 14: Space shuttle part 2

```r
anova(space.shuttle.3,space.shuttle.1)

## Likelihood ratio tests of ordinal regression models
##
## Response: distress
##         Model Resid. df Resid. Dev   Test    Df LR stat.     Pr(Chi)
## 1        temp        20   47.92046
## 2 temp + date        19   37.59413 1 vs 2     1 10.32633 0.001311456
```

Figure 15: Space shuttle part 3

```r
temps=c(67,75)
dates=c(8600,9300)
new=expand.grid(temp=temps,date=dates)
pp=predict(space.shuttle.1,new,type="p")
cbind(new,pp)

##   temp date          0       1-2        3+
## 1   67 8600 0.44659249 0.3679833 0.1854242
## 2   75 8600 0.76360667 0.1825854 0.0538079
## 3   67 9300 0.07472014 0.2306452 0.6946346
## 4   75 9300 0.24428243 0.3933556 0.3623620
```

Figure 16: Predictions 1

```r
new=data.frame(temp=31,date=9524)
pp=predict(space.shuttle.1,new,type="p")
cbind(new,pp)

##     temp date           pp
## 0     31 9524 7.526113e-05
## 1-2   31 9524 3.343053e-04
## 3+    31 9524 9.995904e-01
```

Figure 17: Prediction 2

```
pbc1=read.table("pbc.txt",header=T)
str(pbc1)

## 'data.frame': 312 obs. of  9 variables:
##  $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ days    : int  400 4500 1012 1925 1504 2503 1832 2466 2400 51 ...
##  $ status  : int  2 0 2 2 1 2 0 2 2 2 ...
##  $ drug    : int  1 1 1 1 2 2 2 2 1 2 ...
##  $ age     : int  21464 20617 25594 19994 13918 24201 20284 19379 15526 25772 ...
##  $ edema   : num  1 0 0.5 0.5 0 0 0 0 0 1 ...
##  $ bilirubi: num  14.5 1.1 1.4 1.8 3.4 0.8 1 0.3 3.2 12.6 ...
##  $ albumin : num  2.6 4.14 3.48 2.54 3.53 3.98 4.09 4 3.08 2.74 ...
##  $ prothom : num  12.2 10.6 12 10.3 10.9 11 9.7 11 11 11.5 ...
```

Figure 18: Primary biliary cirrhosis data

```
library(survival)
attach(pbc1)
y=Surv(days,status==2)
head(y,n=20)

##  [1]  400  4500+ 1012  1925  1504+ 2503  1832+ 2466  2400    51  3762
## [12]  304  3577+ 1217  3584  3672+  769   131  4232+ 1356
```

Figure 19: Survival analysis part 1

```r
y.1=coxph(y~drug+age+edema+bilirubi+albumin+prothom)
summary(y.1)

## Call:
## coxph(formula = y ~ drug + age + edema + bilirubi + albumin +
##     prothom)
##
##   n= 312, number of events= 125
##
##                coef  exp(coef)   se(coef)      z Pr(>|z|)
## drug     -1.217e-02  9.879e-01  1.847e-01 -0.066 0.947446
## age       9.036e-05  1.000e+00  2.569e-05  3.517 0.000436 ***
## edema     8.198e-01  2.270e+00  3.104e-01  2.641 0.008262 **
## bilirubi  1.160e-01  1.123e+00  1.507e-02  7.699 1.37e-14 ***
## albumin  -1.210e+00  2.981e-01  2.345e-01 -5.162 2.45e-07 ***
## prothom   2.658e-01  1.304e+00  7.403e-02  3.590 0.000330 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## drug        0.9879     1.0122    0.6879     1.419
## age         1.0001     0.9999    1.0000     1.000
## edema       2.2701     0.4405    1.2355     4.171
## bilirubi    1.1230     0.8904    1.0903     1.157
## albumin     0.2981     3.3544    0.1883     0.472
## prothom     1.3045     0.7666    1.1283     1.508
##
## Concordance= 0.827  (se = 0.02 )
## Likelihood ratio test= 163.5  on 6 df,   p=<2e-16
## Wald test            = 183.7  on 6 df,   p=<2e-16
## Score (logrank) test = 284.7  on 6 df,   p=<2e-16
```

Figure 20: Survival analysis part 2

```r
drugs=c(1,2)
ages=c(15000,20000)
edemas=1
bilirubis=3
albumins=3.5
prothoms=10.5
new=expand.grid(drug=drugs,age=ages,edema=edemas,bilirubi=bilirubis,
  albumin=albumins,prothom=prothoms)
colours=c("red","blue","green","black")
cbind(new,colours)
```

```
##   drug   age edema bilirubi albumin prothom colours
## 1    1 15000     1        3     3.5    10.5     red
## 2    2 15000     1        3     3.5    10.5    blue
## 3    1 20000     1        3     3.5    10.5   green
## 4    2 20000     1        3     3.5    10.5   black
```

Figure 21: Survival analysis part 3

```
pp=survfit(y.1,new)
plot(pp,col=colours)
```

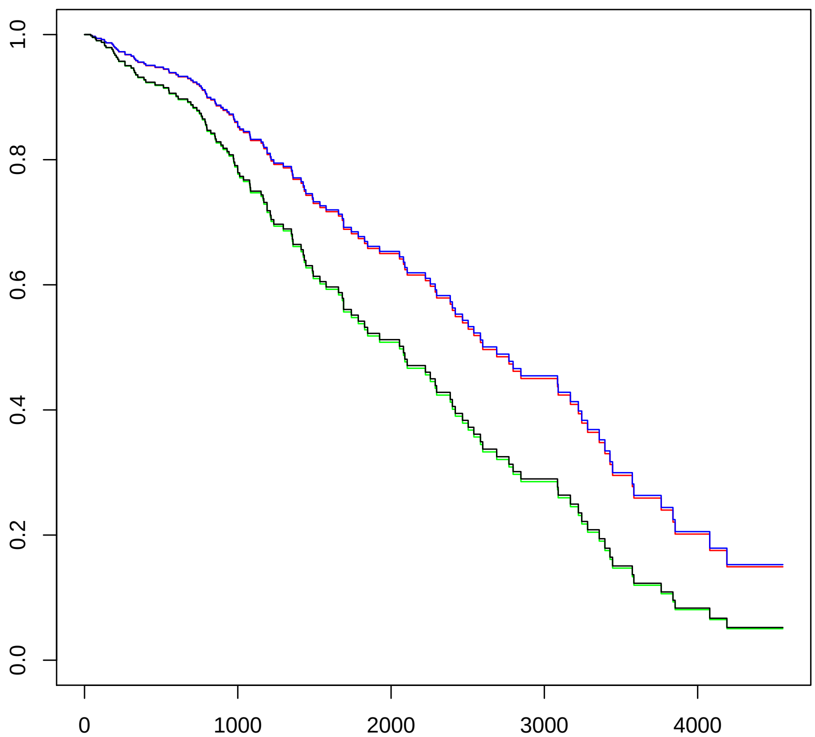


Figure 22: Survival analysis part 4