Code Appendix

Jared Yu

December 7, 2018

```
library(readxl)
baby <- read_excel('baby.xls')
baby
```

```
## # A tibble: 189 x 7
##      age weight smoke pre   hyp   visits birth
##    <dbl>  <dbl> <chr> <chr> <chr>  <dbl> <dbl>
##  1   19.   182. no    no    no        0.    1.
##  2   33.   155. no    no    no        3.    1.
##  3   20.   105. yes   no    no        1.    1.
##  4   21.   108. yes   no    no        2.    1.
##  5   18.   107. yes   no    no        0.    1.
##  6   21.   124. no    no    no        0.    1.
##  7   22.   118. no    no    no        1.    1.
##  8   17.   103. no    no    no        1.    1.
##  9   29.   123. yes   no    no        1.    1.
## 10   26.   113. yes   no    no        0.    1.
## # ... with 179 more rows
```

```
#### table the number of 1's 0's in birth
table(baby$birth)
```

```
##
##   0   1
##  59 130
```

```
130/(59+130)
```

```
## [1] 0.6878307
```

B.1

$\hat{\beta}, s(\hat{\beta}), z-value, p-value$

```
baby.fit = glm(birth~., family=binomial(), data=baby)
summary(baby.fit)
```

```
##
## Call:
## glm(formula = birth ~ ., family = binomial(), data = baby)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2714  -0.8509   0.6034   0.8083   1.8163
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.021488   1.113152  -1.816  0.06937 .
## age          0.059091   0.036965   1.599  0.10992
## weight       0.016086   0.006943   2.317  0.02051 *
## smokeyes    -0.513740   0.349295  -1.471  0.14135
## preyes      -1.798908   0.510014  -3.527  0.00042 ***
## hypyes      -1.772643   0.717756  -2.470  0.01352 *
## visits       0.032113   0.178906   0.179  0.85755
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 234.67  on 188  degrees of freedom
## Residual deviance: 202.15  on 182  degrees of freedom
## AIC: 216.15
##
## Number of Fisher Scoring iterations: 4
```

```
G = 202.15
df = 182
G > qchisq(0.95, df) # fail to rej. H_0
```

```
## [1] FALSE
```

```
1-pchisq(G, df)
```

```
## [1] 0.1459073
```

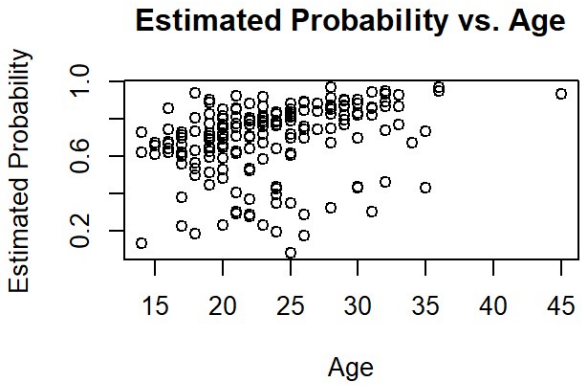
$exp(\hat{\beta})$

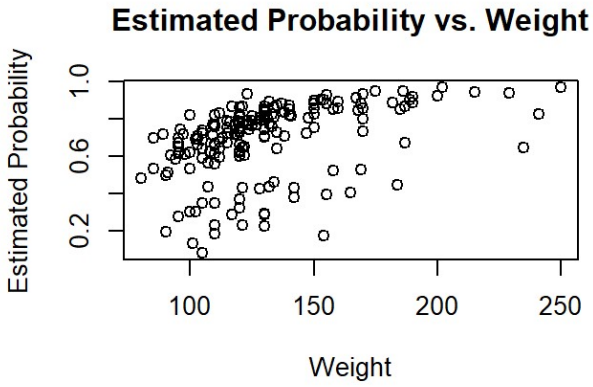
```
for (i in 2:7){ # exp(beta0-beta_6)
  print(exp(baby.fit$coefficients[i]))
}
```

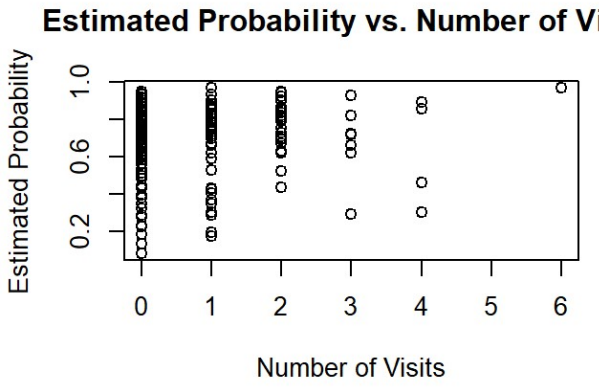
```
##      age
## 1.060872
##   weight
## 1.016216
##  smokeyes
## 0.5982541
##    preyes
## 0.1654794
##    hypyes
## 0.1698834
##   visits
## 1.032634
```

End of B.1

```
par(mfrow=c(2,2))
plot(baby$age, baby.fit$fitted.values, main = 'Estimated Probability vs. Age',
     xlab = 'Age', ylab = 'Estimated Probability')
plot(baby$weight, baby.fit$fitted.values, main = 'Estimated Probability vs. Weight',
     xlab = 'Weight', ylab = 'Estimated Probability')
plot(baby$visits, baby.fit$fitted.values, main = 'Estimated Probability vs. Number of
Visits',
     xlab = 'Number of Visits', ylab = 'Estimated Probability')
```



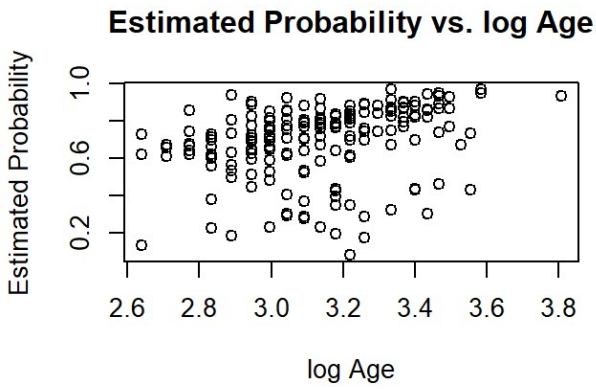


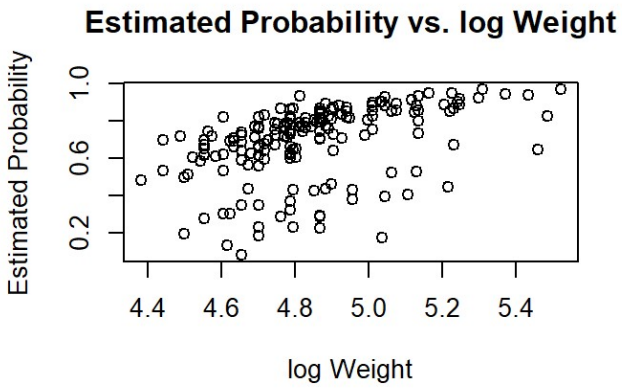


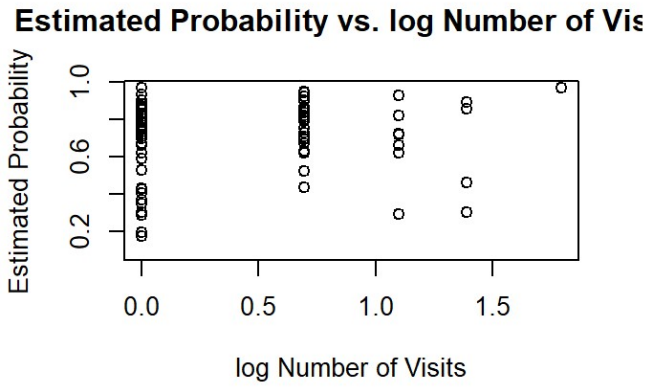

```
dev.off()
```

```
## null device
##           1
```

```
par(mfrow=c(2,2))
plot(log(baby$age), baby.fit$fitted.values, main = 'Estimated Probability vs. log Ag
e',
     xlab = 'log Age', ylab = 'Estimated Probability')
plot(log(baby$weight), baby.fit$fitted.values, main = 'Estimated Probability vs. log W
eight',
     xlab = 'log Weight', ylab = 'Estimated Probability')
plot(log(baby$visits), baby.fit$fitted.values, main = 'Estimated Probability vs. log N
umber of Visits',
     xlab = 'log Number of Visits', ylab = 'Estimated Probability')
```








```
dev.off()
```

```
## null device
##           1
```

```
baby_bin <- baby
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

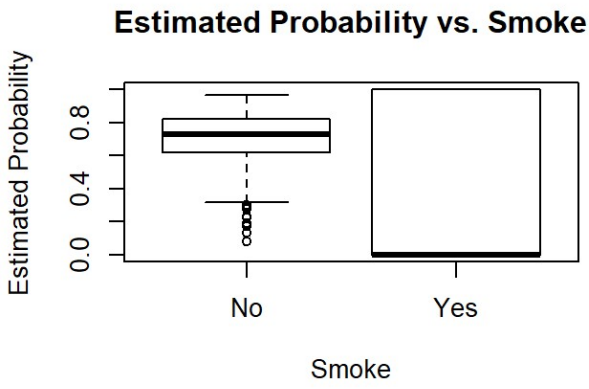
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

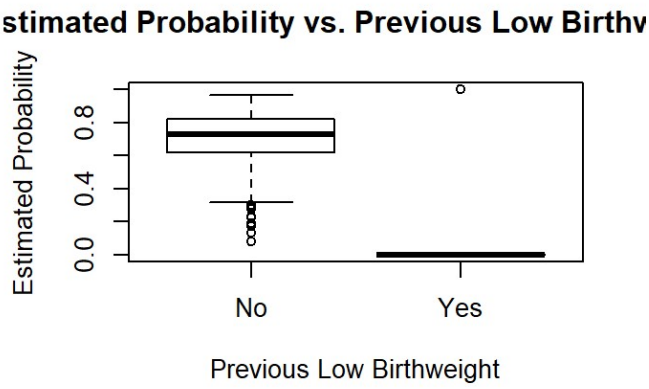
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

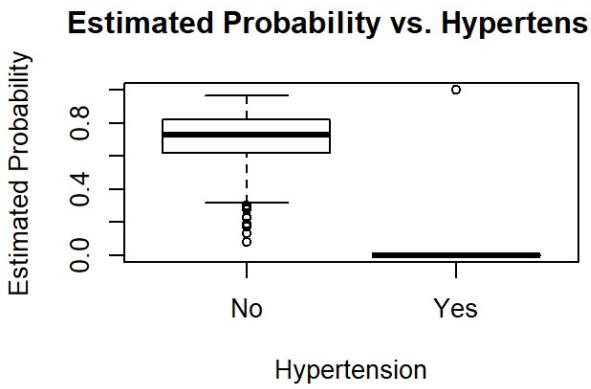
```
baby_bin <- baby_bin %>%
  mutate(smoke = ifelse(smoke == "no",0,1),
         pre = ifelse(pre == "no",0,1),
         hyp = ifelse(hyp == "no",0,1))
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
par(mfrow=c(2,2))
boxplot(baby.fit$fitted.values, baby_bin$smoke, main = 'Estimated Probability vs. Smok
e',
        names = c('No', 'Yes'), xlab = 'Smoke', ylab = 'Estimated Probability')
boxplot(baby.fit$fitted.values, baby_bin$pre, names = c('No', 'Yes'),
        main = 'Estimated Probability vs. Previous Low Birthweight',
        xlab = 'Previous Low Birthweight', ylab = 'Estimated Probability')
boxplot(baby.fit$fitted.values, baby_bin$hyp, main = 'Estimated Probability vs. Hypert
ension',
        names = c('No', 'Yes'), xlab = 'Hypertension', ylab = 'Estimated Probability')
```








```
dev.off()
```

```
## null device
##           1
```

```
baby.fit$coefficients # MLE for beta0 to beta6
```

```
## (Intercept)         age      weight    smokeyes      preyes      hypyes
## -2.02148830  0.05909088  0.01608554 -0.51373966 -1.79890839 -1.77264299
##      visits
##  0.03211329
```

```
summary(baby.fit)$coefficients[,2] # standard error for beta0 to beta6
```

```
## (Intercept)         age      weight    smokeyes      preyes      hypyes
## 1.113152268 0.036964977 0.006942757 0.349295126 0.510014470 0.717755674
##      visits
## 0.178905929
```

```
baby2 = baby # interaction terms
baby2$hyp <- as.factor(baby2$hyp)
baby2$pre <- as.factor(baby2$pre)
inter.fit = glm(birth~age+weight+smoke+pre+hyp+visits+age*weight+weight*hyp+weight*pr
e,
                family = binomial(), data = baby2)
inter.fit
```

```
##
## Call:  glm(formula = birth ~ age + weight + smoke + pre + hyp + visits +
##     age * weight + weight * hyp + weight * pre, family = binomial(),
##     data = baby2)
##
## Coefficients:
##   (Intercept)            age         weight       smokeyes         preyes
##    -3.2742473      0.1165393      0.0258898     -0.5087686     -2.6303649
##        hypyes         visits     age:weight  weight:hypyes  weight:preyes
##    -1.6168753      0.0292054     -0.0004465     -0.0009711      0.0064942
##
## Degrees of Freedom: 188 Total (i.e. Null);  179 Residual
## Null Deviance:       234.7
## Residual Deviance: 202   AIC: 222
```

```
summary(inter.fit)
```

```
##
## Call:
## glm(formula = birth ~ age + weight + smoke + pre + hyp + visits +
##     age * weight + weight * hyp + weight * pre, family = binomial(),
##     data = baby2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2849  -0.8456   0.5938   0.8057   1.9052
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -3.2742473  3.9295522  -0.833    0.405
## age            0.1165393  0.1682394   0.693    0.488
## weight         0.0258898  0.0304761   0.850    0.396
## smokeyes      -0.5087686  0.3545077  -1.435    0.151
## preyes        -2.6303649  2.8680721  -0.917    0.359
## hypyes        -1.6168753  2.6785514  -0.604    0.546
## visits         0.0292054  0.1802465   0.162    0.871
## age:weight    -0.0004465  0.0012718  -0.351    0.726
## weight:hypyes -0.0009711  0.0171863  -0.057    0.955
## weight:preyes  0.0064942  0.0222361   0.292    0.770
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 234.67  on 188  degrees of freedom
## Residual deviance: 201.96  on 179  degrees of freedom
## AIC: 221.96
##
## Number of Fisher Scoring iterations: 4
```

```
inter.fit2 = glm(birth~log(age)+log(weight)+smoke+pre+hyp+visits+ # log interaction te
rms
                   log(age)*log(weight)+log(weight)*hyp+log(weight)*pre,
                family = binomial(), data = baby2)
```

B.2

```
library(MASS) # Stepwise
```

```
##
## Attaching package: 'MASS'
```

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

```
step <- stepAIC(inter.fit, trace = TRUE, direction="backward")
```

```
## Start:  AIC=221.96
## birth ~ age + weight + smoke + pre + hyp + visits + age * weight +
##     weight * hyp + weight * pre
##
##              Df Deviance    AIC
## - weight:hyp  1   201.97 219.97
## - visits      1   201.99 219.99
## - weight:pre  1   202.05 220.05
## - age:weight  1   202.08 220.08
## <none>            201.96 221.96
## - smoke       1   204.01 222.01
##
## Step:  AIC=219.97
## birth ~ age + weight + smoke + pre + hyp + visits + age:weight +
##     weight:pre
##
##              Df Deviance    AIC
## - visits      1   201.99 217.99
## - weight:pre  1   202.06 218.06
## - age:weight  1   202.09 218.09
## <none>            201.97 219.97
## - smoke       1   204.08 220.08
## - hyp         1   208.42 224.42
##
## Step:  AIC=217.99
## birth ~ age + weight + smoke + pre + hyp + age:weight + weight:pre
##
##              Df Deviance    AIC
## - weight:pre  1   202.09 216.09
## - age:weight  1   202.12 216.12
## <none>            201.99 217.99
## - smoke       1   204.16 218.16
## - hyp         1   208.55 222.55
##
## Step:  AIC=216.08
## birth ~ age + weight + smoke + pre + hyp + age:weight
##
##              Df Deviance    AIC
## - age:weight  1   202.19 214.19
## <none>            202.09 216.09
## - smoke       1   204.32 216.32
## - hyp         1   208.74 220.74
## - pre         1   215.65 227.65
##
## Step:  AIC=214.19
## birth ~ age + weight + smoke + pre + hyp
##
##          Df Deviance    AIC
```

```
## <none>        202.19 214.19
## - smoke   1   204.39 214.39
## - age     1   205.09 215.09
## - weight  1   208.42 218.42
## - hyp     1   208.77 218.77
## - pre     1   215.68 225.68
```

```
step$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## birth ~ age + weight + smoke + pre + hyp + visits + age * weight +
##     weight * hyp + weight * pre
##
## Final Model:
## birth ~ age + weight + smoke + pre + hyp
##
##
##           Step Df    Deviance Resid. Df Resid. Dev      AIC
## 1                                   179   201.9624 221.9624
## 2 - weight:hyp  1 0.003177909       180   201.9655 219.9655
## 3     - visits  1 0.028849989       181   201.9944 217.9944
## 4 - weight:pre  1 0.090274230       182   202.0847 216.0847
## 5 - age:weight  1 0.101303736       183   202.1860 214.1860
```

```
test_stat = 202.19-202.15 # new - old
df = 183-182
test_stat > qchisq(0.95, df) # fail to rej. H_0
```

```
## [1] FALSE
```

```
1-pchisq(test_stat, df) # 0.8414806
```

```
## [1] 0.8414806
```

$\hat{\beta}, s(\hat{\beta}), z-value, p-value$

```
summary(step)
```

```
##
## Call:
## glm(formula = birth ~ age + weight + smoke + pre + hyp, family = binomial(),
##     data = baby2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2571  -0.8519   0.6077   0.8016   1.8193
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.031969   1.111157  -1.829 0.067445 .
## age          0.060319   0.036317   1.661 0.096735 .
## weight       0.016154   0.006923   2.333 0.019625 *
## smokeyes    -0.518366   0.348309  -1.488 0.136688
## preyes      -1.794038   0.508841  -3.526 0.000422 ***
## hypyes      -1.782710   0.716698  -2.487 0.012868 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 234.67  on 188  degrees of freedom
## Residual deviance: 202.19  on 183  degrees of freedom
## AIC: 214.19
##
## Number of Fisher Scoring iterations: 4
```

$exp(\beta)$

```
for (i in 1:6){
  print(exp(step$coefficients[i]))
}
```

```
## (Intercept)
##   0.1310772
##      age
## 1.062175
##   weight
## 1.016286
##  smokeyes
## 0.5954926
##    preyes
## 0.1662874
##    hypyes
## 0.1681818
```

End of B.2

B.3

```
step_red <- stepAIC(baby.fit, trace = TRUE, direction="backward")
```

```
## Start:  AIC=216.15
## birth ~ age + weight + smoke + pre + hyp + visits
##
##          Df Deviance    AIC
## - visits  1   202.19 214.19
## <none>        202.15 216.15
## - smoke   1   204.30 216.30
## - age     1   204.83 216.83
## - weight  1   208.26 220.26
## - hyp     1   208.65 220.65
## - pre     1   215.68 227.68
##
## Step:  AIC=214.19
## birth ~ age + weight + smoke + pre + hyp
##
##          Df Deviance    AIC
## <none>        202.19 214.19
## - smoke   1   204.39 214.39
## - age     1   205.09 215.09
## - weight  1   208.42 218.42
## - hyp     1   208.77 218.77
## - pre     1   215.68 225.68
```

```
step_red$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## birth ~ age + weight + smoke + pre + hyp + visits
##
## Final Model:
## birth ~ age + weight + smoke + pre + hyp
##
##
##       Step Df   Deviance Resid. Df Resid. Dev      AIC
## 1                              182   202.1536 216.1536
## 2 - visits  1 0.03233269       183   202.1860 214.1860
```

End of B.3

B.4

```
step2 <- stepAIC(inter.fit2, trace = FALSE,direction="backward")
step2$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## birth ~ log(age) + log(weight) + smoke + pre + hyp + visits +
##     log(age) * log(weight) + log(weight) * hyp + log(weight) *
##     pre
##
## Final Model:
## birth ~ log(age) + log(weight) + smoke + pre + hyp
##
##
##                     Step Df     Deviance Resid. Df Resid. Dev      AIC
## 1                                            179   202.1236 222.1236
## 2      - log(weight):pre  1 0.0001864837       180   202.1238 220.1238
## 3 - log(age):log(weight)  1 0.0001451501       181   202.1239 218.1239
## 4               - visits  1 0.0272845793       182   202.1512 216.1512
## 5      - log(weight):hyp  1 0.0760684548       183   202.2273 214.2273
```

```
summary(step2)
```

```
##
## Call:
## glm(formula = birth ~ log(age) + log(weight) + smoke + pre +
##     hyp, family = binomial(), data = baby2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2269  -0.8527   0.5960   0.8008   1.8659
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -13.5666     4.7957  -2.829 0.004671 **
## log(age)      1.2881     0.8287   1.554 0.120118
## log(weight)   2.2719     0.9189   2.472 0.013426 *
## smokeyes     -0.5004     0.3487  -1.435 0.151315
## preyes       -1.7973     0.5073  -3.543 0.000395 ***
## hypyes       -1.7522     0.6991  -2.506 0.012197 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 234.67  on 188  degrees of freedom
## Residual deviance: 202.23  on 183  degrees of freedom
## AIC: 214.23
##
## Number of Fisher Scoring iterations: 4
```

End of B.4

```
par(mfrow=c(2,2))
res.D = residuals(step, type = 'deviance')
plot(res.D, main = 'Deviance Residuals vs. Index', ylab = 'Deviance Residuals')
range(res.D)
```

```
## [1] -2.257052  1.819297
```

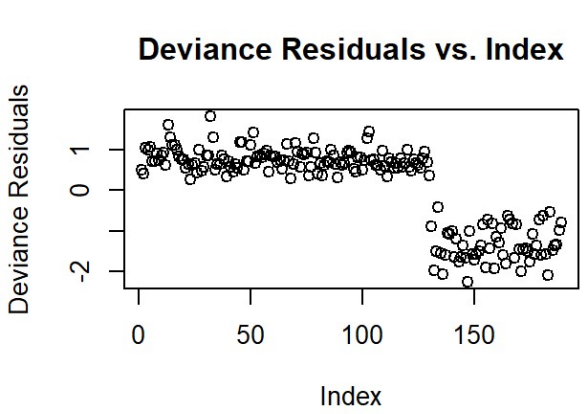
```
res.D.standard = rstudent(step)
plot(res.D.standard, main = 'Standardized Deviance Residuals vs. Index',
     ylab = 'Standardized Deviance Residuals')
range(res.D.standard)
```

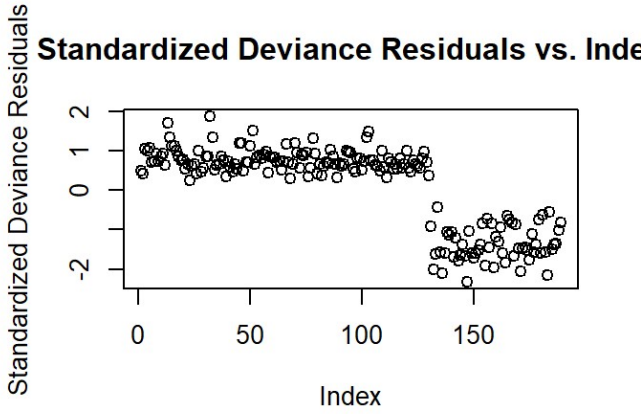
```
## [1] -2.325042  1.872101
```

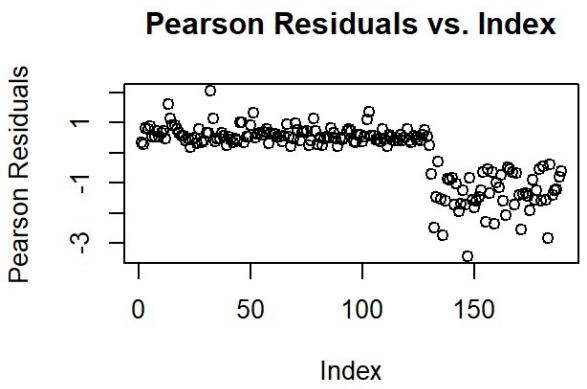
```
res.P = residuals(step, type = 'pearson')
plot(res.P, main = 'Pearson Residuals vs. Index', ylab = 'Pearson Residuals')
range(res.P)
```

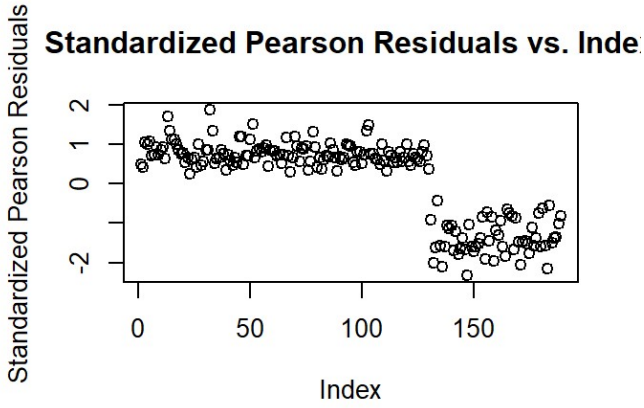
```
## [1] -3.430822  2.057345
```

```
res.P.standard = rstandard(step)
plot(res.D.standard, main = 'Standardized Pearson Residuals vs. Index',
     ylab = 'Standardized Pearson Residuals')
```










```
range(res.P.standard)
```

```
## [1] -2.286726  1.860716
```

```
dev.off()
```

```
## null device
##           1
```

```
table(baby2$birth == 1, inter.fit$fitted.values > 0.5)
```

```
##
##         FALSE TRUE
##   FALSE    21   38
##   TRUE     10  120
```

```
# age
c((step$coefficients[2]) - 1.96*(summary(step)$coefficients[,2][2]),
  (step$coefficients[2]) + 1.96*(summary(step)$coefficients[,2][2]))
```

```
##         age         age
## -0.01086269  0.13149973
```

```
# weight
c((step$coefficients[3]) - 1.96*(summary(step)$coefficients[,2][3]),
  (step$coefficients[3]) + 1.96*(summary(step)$coefficients[,2][3]))
```

```
##      weight      weight
## 0.002585301 0.029723494
```

```
# smoke
c((step$coefficients[4]) - 1.96*(summary(step)$coefficients[,2][4]),
  (step$coefficients[4]) + 1.96*(summary(step)$coefficients[,2][4]))
```

```
##   smokeyes   smokeyes
## -1.2010521  0.1643193
```

```
# pre
c((step$coefficients[5]) - 1.96*(summary(step)$coefficients[,2][5]),
  (step$coefficients[5]) + 1.96*(summary(step)$coefficients[,2][5]))
```

```
##    preyes    preyes
## -2.791367 -0.796709
```

```
# hyp
c((step$coefficients[6]) - 1.96*(summary(step)$coefficients[,2][6]),
  (step$coefficients[6]) + 1.96*(summary(step)$coefficients[,2][6]))
```

```
##     hypyes     hypyes
## -3.1874385 -0.3779811
```

```
ischemic <- read_excel('ischemic.xlsx')
ischemic
```

```
## # A tibble: 788 x 9
##      cost   age gender inter drugs complications comorbidities duration
##     <dbl> <dbl>  <dbl> <dbl> <dbl>         <dbl>         <dbl>    <dbl>
##  1   179.   63.     0.    2.    1.            0.            3.     300.
##  2   319.   59.     0.    2.    0.            0.            0.     120.
##  3  9311.   62.     0.   17.    0.            0.            5.     353.
##  4   281.   60.     1.    9.    0.            0.            2.     332.
##  5 18727.   55.     0.    5.    2.            0.            0.      18.
##  6   453.   66.     0.    1.    0.            0.            4.     296.
##  7   323.   64.     1.    2.    0.            0.            1.     247.
##  8  3874.   45.     1.    3.    0.            0.            1.      82.
##  9  3244.   68.     0.    6.    2.            0.            4.     334.
## 10   226.   64.     1.    3.    0.            0.            0.      85.
## # ... with 778 more rows, and 1 more variable: visits <dbl>
```

```
mean(ischemic$visits)
```

```
## [1] 3.425127
```

```
var(ischemic$visits)
```

```
## [1] 6.956267
```

B.5

$\hat{\beta}, s(\hat{\beta}), z - value, p - value$

```
fit.poisson <- glm(visits~., data = ischemic, family = poisson())
summary(fit.poisson)
```

```
##
## Call:
## glm(formula = visits ~ ., family = poisson(), data = ischemic)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6851  -1.0341  -0.2372   0.5845   5.7602
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    4.994e-01  1.761e-01   2.837  0.00456 **
## cost           1.495e-05  2.855e-06   5.237 1.63e-07 ***
## age            6.724e-03  2.967e-03   2.266  0.02346 *
## gender         1.819e-01  4.400e-02   4.135 3.55e-05 ***
## inter          1.007e-02  3.808e-03   2.646  0.00816 **
## drugs          1.932e-01  1.268e-02  15.234  < 2e-16 ***
## complications  6.125e-02  5.995e-02   1.022  0.30689
## comorbidities -8.999e-04  3.685e-03  -0.244  0.80708
## duration       3.529e-04  1.899e-04   1.859  0.06308 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1485.0  on 787  degrees of freedom
## Residual deviance: 1043.6  on 779  degrees of freedom
## AIC: 3271
##
## Number of Fisher Scoring iterations: 5
```

```
G=1043.6
df=779
G > qchisq(0.95, df) # fail to rej. H_0
```

```
## [1] TRUE
```

```
1-pchisq(G, df)
```

```
## [1] 5.893235e-10
```

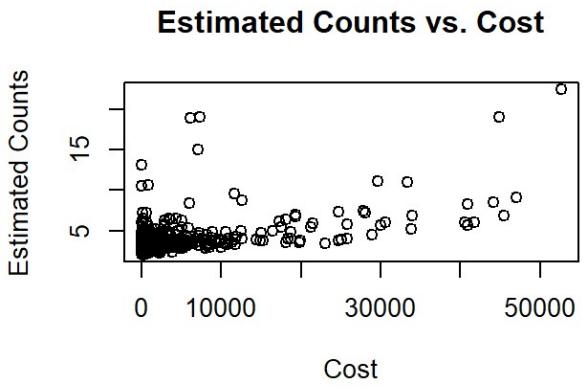
$exp(\hat{\beta})$

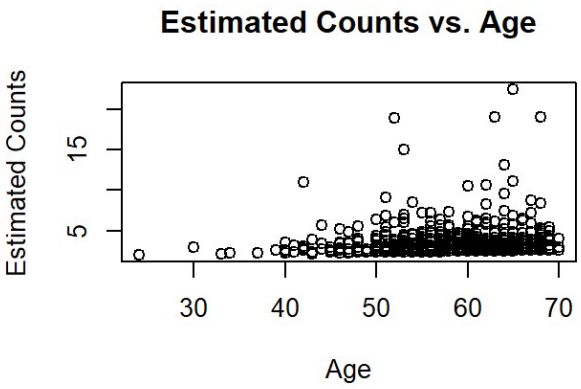
```
for (i in 1:9){
  print(exp(fit.poisson$coefficients[i]))
}
```

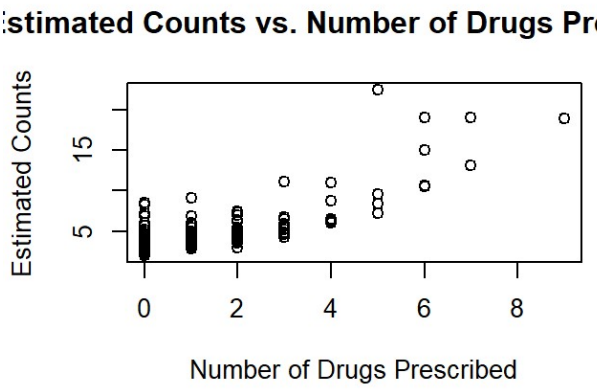
```
## (Intercept)
##    1.647808
##     cost
## 1.000015
##      age
## 1.006747
##   gender
## 1.199518
##    inter
## 1.010126
##   drugs
## 1.21317
## complications
##       1.06317
## comorbidities
##     0.9991005
## duration
## 1.000353
```

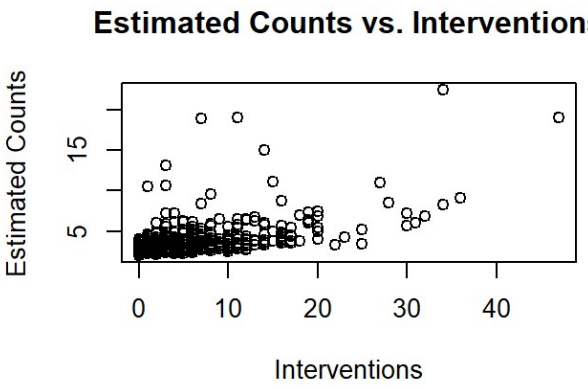
End of B.5

```
par(mfrow=c(2,2))
plot(ischemic$cost, fit.poisson$fitted.values, main = 'Estimated Counts vs. Cost',
     xlab = 'Cost', ylab = 'Estimated Counts')
plot(ischemic$age, fit.poisson$fitted.values, main = 'Estimated Counts vs. Age',
     xlab = 'Age', ylab = 'Estimated Counts')
plot(ischemic$drugs, fit.poisson$fitted.values, main = 'Estimated Counts vs. Number o
f Drugs Prescribed',
     xlab = 'Number of Drugs Prescribed', ylab = 'Estimated Counts')
plot(ischemic$inter, fit.poisson$fitted.values, main = 'Estimated Counts vs. Intervent
ions',
     xlab = 'Interventions', ylab = 'Estimated Counts')
```





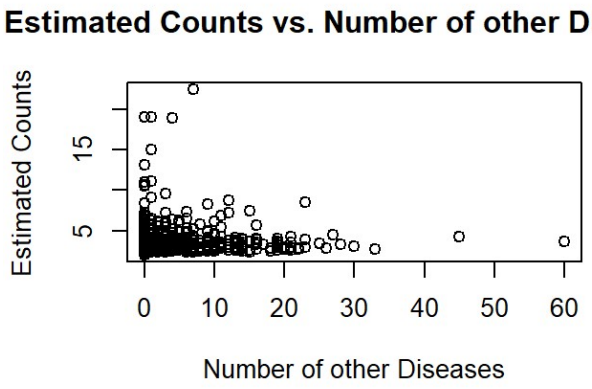


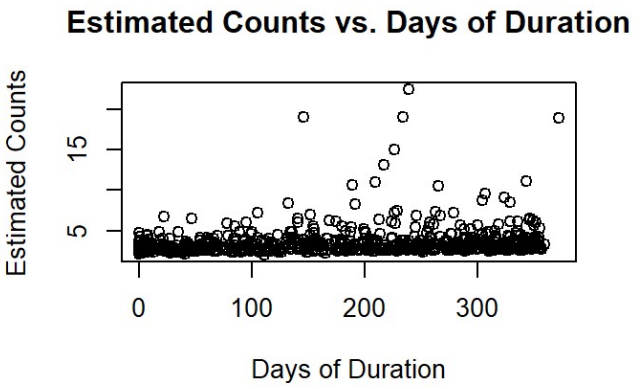


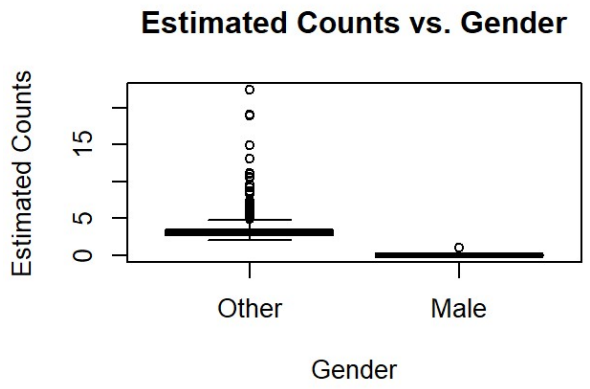

```
dev.off()
```

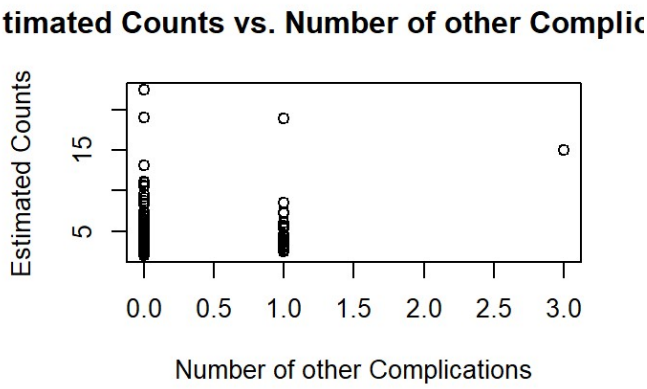
```
## null device
##           1
```

```
par(mfrow=c(2,2))
plot(ischemic$comorbidities, fit.poisson$fitted.values, main = 'Estimated Counts vs. N
umber of other Diseases',
     xlab = 'Number of other Diseases', ylab = 'Estimated Counts')
plot(ischemic$duration, fit.poisson$fitted.values, main = 'Estimated Counts vs. Days o
f Duration',
     xlab = 'Days of Duration', ylab = 'Estimated Counts')
boxplot(fit.poisson$fitted.values, ischemic$gender, main = 'Estimated Counts vs. Gende
r',
        xlab = 'Gender', ylab = 'Estimated Counts', names = c('Other', 'Male'))
plot(ischemic$complications, fit.poisson$fitted.values, main = 'Estimated Counts vs. N
umber of other Complications',
     xlab = 'Number of other Complications', ylab = 'Estimated Counts')
```










```
dev.off()
```

```
## null device
##           1
```

B.6

```
ischemic2 = ischemic
for (i in setdiff(1:9,3)){
    ischemic2[i] = sqrt(ischemic2[i])
}
options(warn=-1)
sqrt_poisson <- glm(visits~., data = ischemic2, family = poisson()) # warnings
options(warn=0)
summary(sqrt_poisson)
```

```
##
## Call:
## glm(formula = visits ~ ., family = poisson(), data = ischemic2)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.91969  -0.32392   0.01945   0.27983   1.35175
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.3730641  0.4777439  -0.781  0.43487
## cost           0.0025069  0.0009553   2.624  0.00868 **
## age            0.0868825  0.0624003   1.392  0.16382
## gender         0.0811436  0.0639617   1.269  0.20457
## inter         -0.0012630  0.0330922  -0.038  0.96955
## drugs          0.2388473  0.0419716   5.691 1.27e-08 ***
## complications  0.0624833  0.1080390   0.578  0.56303
## comorbidities -0.0099807  0.0240665  -0.415  0.67835
## duration       0.0042073  0.0056873   0.740  0.45944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 307.66  on 787  degrees of freedom
## Residual deviance: 239.88  on 779  degrees of freedom
## AIC: Inf
##
## Number of Fisher Scoring iterations: 4
```

End of B.6

B.7

```
step_poi <- stepAIC(fit.poisson, direction = 'backward', trace=TRUE)
```

```
## Start:  AIC=3271.03
## visits ~ cost + age + gender + inter + drugs + complications +
##     comorbidities + duration
##
##                 Df Deviance    AIC
## - comorbidities  1   1043.7 3269.1
## - complications  1   1044.6 3270.1
## <none>               1043.6 3271.0
## - duration       1   1047.1 3272.5
## - age            1   1048.8 3274.2
## - inter          1   1050.3 3275.8
## - gender         1   1060.2 3285.6
## - cost           1   1070.2 3295.6
## - drugs          1   1238.5 3463.9
##
## Step:  AIC=3269.09
## visits ~ cost + age + gender + inter + drugs + complications +
##     duration
##
##                 Df Deviance    AIC
## - complications  1   1044.7 3268.1
## <none>               1043.7 3269.1
## - duration       1   1047.5 3270.9
## - age            1   1048.8 3272.2
## - inter          1   1050.3 3273.8
## - gender         1   1060.5 3283.9
## - cost           1   1070.2 3293.6
## - drugs          1   1245.3 3468.7
##
## Step:  AIC=3268.1
## visits ~ cost + age + gender + inter + drugs + duration
##
##            Df Deviance    AIC
## <none>          1044.7 3268.1
## - duration  1   1048.9 3270.3
## - age       1   1049.4 3270.8
## - inter     1   1051.7 3273.2
## - gender    1   1062.1 3283.6
## - cost      1   1071.4 3292.8
## - drugs     1   1254.0 3475.4
```

```
step_poi$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## visits ~ cost + age + gender + inter + drugs + complications +
##     comorbidities + duration
##
## Final Model:
## visits ~ cost + age + gender + inter + drugs + duration
##
##
##               Step Df   Deviance Resid. Df Resid. Dev      AIC
## 1                                      779   1043.619 3271.035
## 2 - comorbidities  1 0.05990609       780   1043.679 3269.095
## 3 - complications  1 1.00621343       781   1044.686 3268.101
```

```
test_stat=1044.7-1043.6
df=781-779
test_stat>qchisq(0.95,df)
```

```
## [1] FALSE
```

```
1-pchisq(test_stat, df)
```

```
## [1] 0.5769498
```

$\hat{\beta}, s(\hat{\beta}), z-value, p-value$

```
summary(step_poi)
```

```
##
## Call:
## glm(formula = visits ~ cost + age + gender + inter + drugs +
##     duration, family = poisson(), data = ischemic)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6057  -1.0366  -0.2380   0.5763   5.7457
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 5.208e-01  1.745e-01   2.986  0.00283 **
## cost        1.493e-05  2.844e-06   5.251 1.52e-07 ***
## age         6.334e-03  2.938e-03   2.156  0.03111 *
## gender      1.857e-01  4.379e-02   4.241 2.23e-05 ***
## inter       1.025e-02  3.781e-03   2.710  0.00673 **
## drugs       1.963e-01  1.221e-02  16.067  < 2e-16 ***
## duration    3.453e-04  1.686e-04   2.048  0.04053 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1485.0  on 787  degrees of freedom
## Residual deviance: 1044.7  on 781  degrees of freedom
## AIC: 3268.1
##
## Number of Fisher Scoring iterations: 5
```

$exp(\beta)$

```
for (i in 1:7){
  print(exp(step_poi$coefficients[i]))
}
```

```
## (Intercept)
##    1.683442
##     cost
## 1.000015
##      age
## 1.006354
##   gender
## 1.204078
##  inter
## 1.0103
##    drugs
## 1.216838
## duration
## 1.000345
```

End of B.7

```
par(mfrow=c(2,2))
res.D = residuals(step_poi, type = 'deviance')
plot(res.D, main = 'Deviance Residuals vs. Index', ylab = 'Deviance Residuals')
range(res.D)
```

```
## [1] -2.605684  5.745722
```

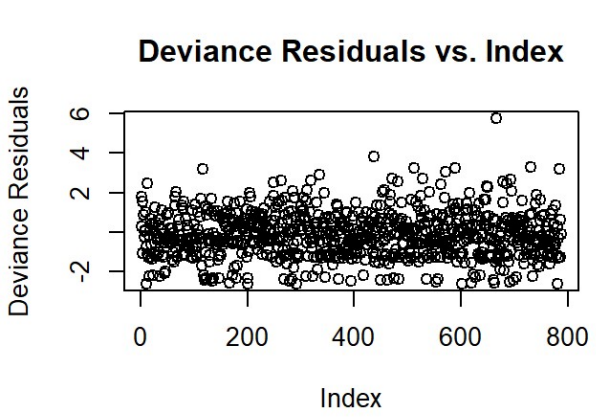
```
res.D.standard = rstudent(step_poi)
plot(res.D.standard, main = 'Standardized Deviance Residuals vs. Index',
     ylab = 'Standardized Deviance Residuals')
range(res.D.standard)
```

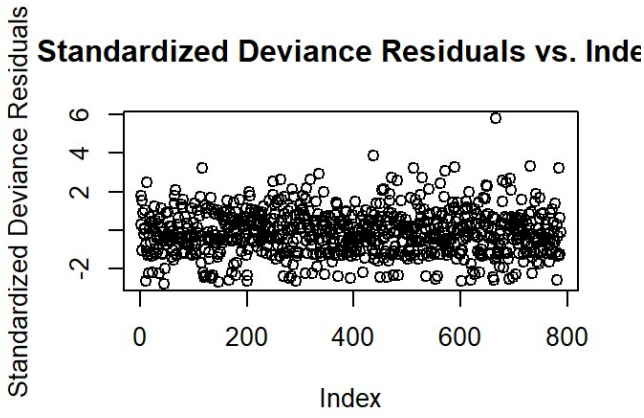
```
## [1] -2.769384  5.785878
```

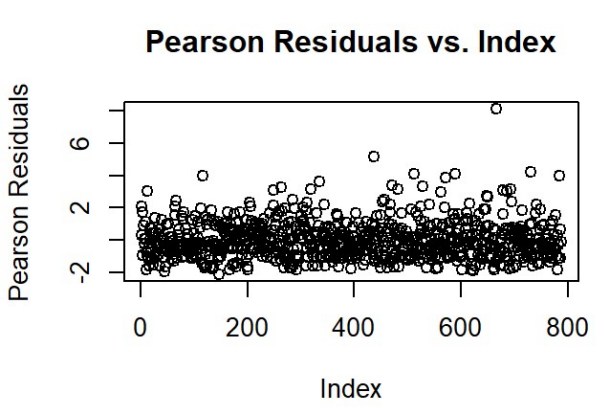
```
res.P = residuals(step_poi, type = 'pearson')
plot(res.P, main = 'Pearson Residuals vs. Index', ylab = 'Pearson Residuals')
range(res.P)
```

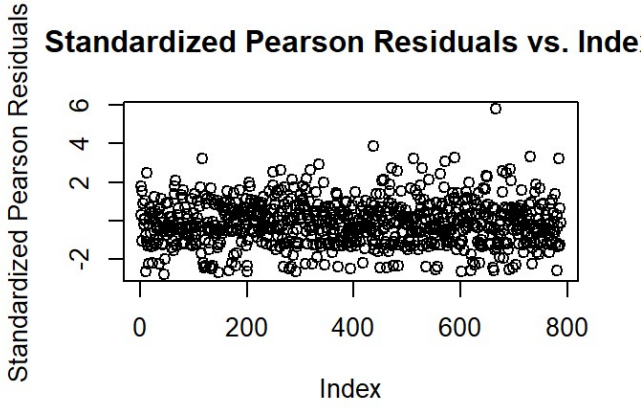
```
## [1] -2.117950  8.119315
```

```
res.P.standard = rstandard(step_poi)
plot(res.D.standard, main = 'Standardized Pearson Residuals vs. Index',
     ylab = 'Standardized Pearson Residuals')
```










```
range(res.P.standard)
```

```
## [1] -2.879518  5.765866
```

```
dev.off()
```

```
## null device
##           1
```

```
res.D_pos = subset(res.D, res.D > 0)
length(subset(res.D_pos, res.D_pos > 3.5)) / length(res.D_pos)
```

```
## [1] 0.005730659
```

```
res.D.standard_pos = subset(res.D.standard, res.D.standard > 0)
length(subset(res.D.standard_pos, res.D.standard_pos > 3.5)) / length(res.D.standard_p
os)
```

```
## [1] 0.005730659
```

```
res.P_pos = subset(res.P, res.P > 0)
length(subset(res.P_pos, res.P_pos > 3.5)) / length(res.P_pos)
```

```
## [1] 0.02578797
```

```
res.P.standard_pos = subset(res.P.standard, res.P.standard > 0)
length(subset(res.P.standard_pos, res.P.standard_pos > 3.5)) / length(res.P.standard_p
os)
```

```
## [1] 0.005730659
```

```
### cost
c((step_poi$coefficients[2]) - 1.96*(summary(step_poi)$coefficients[,2][2]),
  (step_poi$coefficients[2]) + 1.96*(summary(step_poi)$coefficients[,2][2]))
```

```
##         cost         cost
## 9.357017e-06 2.050373e-05
```

```
### age
c((step_poi$coefficients[3]) - 1.96*(summary(step_poi)$coefficients[,2][3]),
  (step_poi$coefficients[3]) + 1.96*(summary(step_poi)$coefficients[,2][3]))
```

```
##          age          age
## 0.0005747559 0.0120932510
```

```
### gender
c((step_poi$coefficients[4]) - 1.96*(summary(step_poi)$coefficients[,2][4]),
  (step_poi$coefficients[4]) + 1.96*(summary(step_poi)$coefficients[,2][4]))
```

```
##     gender     gender
## 0.09988022 0.27154769
```

```
### inter
c((step_poi$coefficients[5]) - 1.96*(summary(step_poi)$coefficients[,2][5]),
  (step_poi$coefficients[5]) + 1.96*(summary(step_poi)$coefficients[,2][5]))
```

```
##       inter       inter
## 0.002836101 0.017658537
```

```
### drugs
c((step_poi$coefficients[6]) - 1.96*(summary(step_poi)$coefficients[,2][6]),
  (step_poi$coefficients[6]) + 1.96*(summary(step_poi)$coefficients[,2][6]))
```

```
##     drugs     drugs
## 0.1723151 0.2201964
```

```
### duration
c((step_poi$coefficients[7]) - 1.96*(summary(step_poi)$coefficients[,2][7]),
  (step_poi$coefficients[7]) + 1.96*(summary(step_poi)$coefficients[,2][7]))
```

```
##     duration     duration
## 1.488865e-05 6.756963e-04
```