PUB490Z HW4

Aaron Tsui

Sleep Trouble

a

```
load("C:/Users/aaron/Documents/nhanes.samp.adult.Rdata")
pa <- glm(nhanes.samp.adult$SleepTrouble ~ nhanes.samp.adult$Gender + nhanes.samp.adult$Age + nhanes.sa
    family = binomial(link = "logit"))
summary(pa)
```

```
##
## Call:
## glm(formula = nhanes.samp.adult$SleepTrouble ~ nhanes.samp.adult$Gender +
##     nhanes.samp.adult$Age + nhanes.samp.adult$Education + nhanes.samp.adult$Poverty +
##     nhanes.samp.adult$Work + nhanes.samp.adult$Depressed, family = binomial(link = "logit"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3723  -0.8217  -0.5746   0.8466   1.9850
##
## Coefficients:
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             -1.941587   2.132318  -0.911  0.36253
## nhanes.samp.adult$Gendermale            -0.730285   0.464476  -1.572  0.11589
## nhanes.samp.adult$Age                   -0.003025   0.016652  -0.182  0.85584
## nhanes.samp.adult$Education9 - 11th Grade  0.927768   1.622414   0.572  0.56743
## nhanes.samp.adult$EducationHigh School     0.970251   1.604930   0.605  0.54548
## nhanes.samp.adult$EducationSome College    0.671384   1.609887   0.417  0.67665
## nhanes.samp.adult$EducationCollege Grad    0.237586   1.636293   0.145  0.88455
## nhanes.samp.adult$Poverty                  0.332506   0.161364   2.061  0.03934
## nhanes.samp.adult$WorkNotWorking          -0.373401   1.356750  -0.275  0.78315
## nhanes.samp.adult$WorkWorking             -0.299420   1.287690  -0.233  0.81613
## nhanes.samp.adult$DepressedSeveral        -0.005107   0.628885  -0.008  0.99352
## nhanes.samp.adult$DepressedMost            3.606329   1.201712   3.001  0.00269
##
## (Intercept)
## nhanes.samp.adult$Gendermale
## nhanes.samp.adult$Age
## nhanes.samp.adult$Education9 - 11th Grade
## nhanes.samp.adult$EducationHigh School
## nhanes.samp.adult$EducationSome College
## nhanes.samp.adult$EducationCollege Grad
## nhanes.samp.adult$Poverty                 *
## nhanes.samp.adult$WorkNotWorking
```

```
## nhanes.samp.adult$WorkWorking
## nhanes.samp.adult$DepressedSeveral
## nhanes.samp.adult$DepressedMost           **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 162.80  on 128  degrees of freedom
## Residual deviance: 134.58  on 117  degrees of freedom
##   (23 observations deleted due to missingness)
## AIC: 158.58
##
## Number of Fisher Scoring iterations: 5
```

b

```
nhanes.samp.adult.2 <- subset(nhanes.samp.adult,
  !(is.na(Depressed)))

pb1 <- glm(SleepTrouble ~ Gender + Age + Education + Poverty + Work + Depressed,
    family = binomial(link = "logit"), data = nhanes.samp.adult.2)
pb2 <- glm(SleepTrouble ~ Gender + Age + Education + Poverty + Work,
    family = binomial(link = "logit"), data = nhanes.samp.adult.2)

anova(pb1, pb2, test = "LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: SleepTrouble ~ Gender + Age + Education + Poverty + Work + Depressed
## Model 2: SleepTrouble ~ Gender + Age + Education + Poverty + Work
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       117     134.58
## 2       119     152.10 -2  -17.523 0.0001567 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With a very low P-value of 0.0001567, we reject the null hypothesis that Depressed has no effect on SleepTrouble. We can conclude that Depressed is a statistically significant predictor of SleepTrouble.

c

```
step(pb1)
```

```
## Start:  AIC=158.58
## SleepTrouble ~ Gender + Age + Education + Poverty + Work + Depressed
##
##             Df Deviance    AIC
## - Education  4   136.62 152.62
## - Work       2   134.65 154.65
## - Age        1   134.61 156.61
## <none>           134.58 158.58
```

```
## - Gender     1   137.14 159.14
## - Poverty    1   139.17 161.17
## - Depressed  2   152.10 172.10
##
## Step:  AIC=152.62
## SleepTrouble ~ Gender + Age + Poverty + Work + Depressed
##
##             Df Deviance    AIC
## - Work       2   136.70 148.70
## - Age        1   136.63 150.63
## <none>           136.62 152.62
## - Gender     1   139.06 153.06
## - Poverty    1   140.11 154.11
## - Depressed  2   154.79 166.79
##
## Step:  AIC=148.7
## SleepTrouble ~ Gender + Age + Poverty + Depressed
##
##             Df Deviance    AIC
## - Age        1   136.75 146.75
## <none>           136.70 148.70
## - Gender     1   139.09 149.09
## - Poverty    1   140.59 150.59
## - Depressed  2   154.95 162.95
##
## Step:  AIC=146.75
## SleepTrouble ~ Gender + Poverty + Depressed
##
##             Df Deviance    AIC
## <none>           136.75 146.75
## - Gender     1   139.24 147.24
## - Poverty    1   140.60 148.60
## - Depressed  2   155.16 161.16
##
## Call:  glm(formula = SleepTrouble ~ Gender + Poverty + Depressed, family = binomial(link = "logit"),
##     data = nhanes.samp.adult.2)
##
## Coefficients:
##      (Intercept)        Gendermale           Poverty  DepressedSeveral
##         -1.55514           -0.69770           0.25987           -0.03678
##    DepressedMost
##          3.45057
##
## Degrees of Freedom: 128 Total (i.e. Null);  124 Residual
##   (4 observations deleted due to missingness)
## Null Deviance:       162.8
## Residual Deviance: 136.7     AIC: 146.7
```

d

Our regression model fits the data better and better as we remove variables from the regression equation because our AIC values decrease as we remove variables from every additional regression model.

e

```
nhanes.samp.adult.2$Depressed <- relevel(nhanes.samp.adult.2$Depressed, ref = "Most")
summary(pb1)
```

```
##
## Call:
## glm(formula = SleepTrouble ~ Gender + Age + Education + Poverty +
##     Work + Depressed, family = binomial(link = "logit"), data = nhanes.samp.adult.2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3723  -0.8217  -0.5746   0.8466   1.9850
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -1.941587   2.132318  -0.911  0.36253
## Gendermale              -0.730285   0.464476  -1.572  0.11589
## Age                     -0.003025   0.016652  -0.182  0.85584
## Education9 - 11th Grade  0.927768   1.622414   0.572  0.56743
## EducationHigh School     0.970251   1.604930   0.605  0.54548
## EducationSome College    0.671384   1.609887   0.417  0.67665
## EducationCollege Grad    0.237586   1.636293   0.145  0.88455
## Poverty                  0.332506   0.161364   2.061  0.03934 *
## WorkNotWorking          -0.373401   1.356750  -0.275  0.78315
## WorkWorking             -0.299420   1.287690  -0.233  0.81613
## DepressedSeveral        -0.005107   0.628885  -0.008  0.99352
## DepressedMost            3.606329   1.201712   3.001  0.00269 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 162.80  on 128  degrees of freedom
## Residual deviance: 134.58  on 117  degrees of freedom
##   (4 observations deleted due to missingness)
## AIC: 158.58
##
## Number of Fisher Scoring iterations: 5
```

The Depressed coefficient values do not change. In terms of odds ratios, SleepTrouble decreases by 0.005107 if the person is Several level of Depressed, but increases by 3.606329 if the person is Most level of Depressed.

f

```
## Warning: package 'caret' was built under R version 4.0.5
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
## Warning: The `i` argument of ``[`()` can't be a matrix as of tibble 3.0.0.
## Convert to a vector.
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
## Warning in mean.default(acc.est.5): argument is not numeric or logical:
## returning NA
```

```
## [1] NA
```

The accuracy of the logistic regression model in part c is good.

ICU: Probability of Survival

a

```
load("C:/Users/aaron/Documents/icu.Rdata")
```

i

```
summary(icu$vital.status)
```

```
## lived  died
##   160    40
```

```
summary(icu$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16.00   46.75   63.00   57.55   72.00   92.00
```

20% (40 of 200) of our study population died before they got to the hospital, whilst the median age for our study population is 57.55. The youngest was 16 and the oldest was 92.

ii

```
lm(age ~ vital.status, data = icu)
```

```
##
## Call:
## lm(formula = age ~ vital.status, data = icu)
##
## Coefficients:
##      (Intercept)  vital.statusdied
##           55.650             9.475
```

```
lm(sys ~ vital.status, data = icu)
```

```
##
## Call:
## lm(formula = sys ~ vital.status, data = icu)
##
## Coefficients:
##      (Intercept)  vital.statusdied
##           135.64            -16.82
```

People who died before they got to the ICU were on average 9.475 years older and had 16.82 less systolic blood pressure. Those who were still living when they got to the ICU were on average 55.65 years of age and had 135.64 systolic blood pressure.

iii

```
lm(service ~ vital.status, data = icu)
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a factor
## response will be ignored
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

```
##
## Call:
## lm(formula = service ~ vital.status, data = icu)
##
## Coefficients:
##      (Intercept)  vital.statusdied
##           1.5812           -0.2312
```

```
lm(type ~ vital.status, data = icu)
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a factor
## response will be ignored
```

```
## Warning in model.response(mf, "numeric"): '-' not meaningful for factors
```

```
##
## Call:
## lm(formula = type ~ vital.status, data = icu)
##
## Coefficients:
##      (Intercept)  vital.statusdied
##           1.6812            0.2688
```

People who died before they got to the ICU were on average .2312 more medical than surgical and 0.2688 more emergency type of admission than elective type of admission. Those who were still living when they got to the ICU were on average .0812 (1.5812 - ((1+2) / 2)) more surgical than medical and .1812 (1.6812 - ((1+2) / 2)) more emergency than elective.

b

i

```
lm(vital.status ~ age + gender + race, data = icu)
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a factor
## response will be ignored
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

```
##
## Call:
## lm(formula = vital.status ~ age + gender + race, data = icu)
##
## Coefficients:
##  (Intercept)           age  genderfemale     raceblack     raceother
##    0.9968224     0.0036301     0.0009541    -0.1020587     0.0314530
```

Race seems to be associated with vital.status, since raceblack seems to affect vital.status significantly more than the other demographic types. I suspect gender and age is not associated with vital.status at all due to their extremely low effect on vital.status.

ii

```
lm(vital.status ~ PO2 + PH + PCO2 + bicarb, data = icu)
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a factor
## response will be ignored
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

```
##
## Call:
## lm(formula = vital.status ~ PO2 + PH + PCO2 + bicarb, data = icu)
##
## Coefficients:
## (Intercept)      PO2< 60     PH> 7.25     PCO2< 45   bicarb< 18
##     1.18534      0.11120      0.08561     -0.06878      0.09439
```

Yes, all of the ABG test measures are associated with vital.status, where PO2 being less than 60 affects vital.status even more than raceblack.

iii

```
lm(vital.status ~ cancer + renal + infect.prob + fracture, data = icu)
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a factor
## response will be ignored
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

```
##
## Call:
## lm(formula = vital.status ~ cancer + renal + infect.prob + fracture,
##     data = icu)
##
## Coefficients:
##    (Intercept)       canceryes        renalyes  infect.probyes     fractureyes
##        1.11914         0.02260         0.22208         0.13474         0.01216
```

History of chronic renal failure and probable infection are the factors the most associated with ICU mortality. Cancer-related admission and presence of a bone fracture both have associations but are far weaker associations than the other 2 variables.

c

i

```
lm(vital.status ~ race + PO2 + bicarb + renal + infect.prob, data = icu)
```

```
## Warning in model.response(mf, "numeric"): using type = "numeric" with a factor
## response will be ignored
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

```
##
## Call:
## lm(formula = vital.status ~ race + PO2 + bicarb + renal + infect.prob,
##     data = icu)
##
## Coefficients:
##    (Intercept)       raceblack       raceother         PO2< 60      bicarb< 18
##        1.13249         -0.12973         -0.01688         0.05313          0.05090
##       renalyes  infect.probyes
##        0.21998          0.11694
```

vital.status = 1.13249 -0.12973*raceblack -0.01688*raceother +0.05313*(PO2<60) +0.05090*(bicarb<18) +0.21998*renalyes +0.11694*infect.probyes

ii

Relative odds of mortality for a patient whom infection was probable is 11.694% higher than relative odds of mortality for a patient whom infection was not probable. The former is 24.943% chances of mortality and the latter is 13.249% chances of mortality.