Quinn Knudsen

Homework 9

Topic: **Categorical Analysis**

1)

> glmOut <- glm(vs ~ gear+hp,data=mtcars, family=binomial())

> summary(glmOut)

Call:

glm(formula = vs ~ gear + hp, family = binomial(), data = mtcars)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.76095 -0.20263 -0.00889 0.38030 1.37305

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 13.43752 7.18161 1.871 0.0613 .

gear -0.96825 1.12809 -0.858 0.3907

hp -0.08005 0.03261 -2.455 0.0141 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 43.860 on 31 degrees of freedom

Residual deviance: 16.013 on 29 degrees of freedom

AIC: 22.013

Number of Fisher Scoring iterations: 7

> anova(glmOut, test="Chisq")

Analysis of Deviance Table

Model: binomial, link: logit

Response: vs

Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Pr(>Chi)

NULL 31 43.860

gear 1 1.3656 30 42.495 0.2426

hp 1 26.4814 29 16.013 2.661e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> exp(coef(glmOut))

(Intercept) gear hp

6.852403e+05 3.797461e-01 9.230734e-01

The results here indicate that hp (-0.08,p<.05) but not gear (ns) predict engine cylinder shape. We can reject the null that horsepower doesn’t differ in the type of engine cylinder shape. There is a negative relationship so that lower horsepower cars predict a straight engine shape.

5)

library(BaylorEdPsych)

> #library(remotes)

> #install\_version("BaylorEdPsych", "0.5")

> PseudoR2(glmOut)

McFadden Adj.McFadden Cox.Snell Nagelkerke McKelvey.Zavoina Effron Count

0.6349042 0.4525061 0.5811397 0.7789526 0.8972195 0.6445327 0.8125000

Adj.Count AIC Corrected.AIC

0.5714286 22.0131402 22.8702830

The Nagelkerke value here is strong at .78.

6)

> glmOut <- glm(vote ~ age + statusquo, data = ChileYN, family=binomial())

> summary(glmOut)

Call:

glm(formula = vote ~ age + statusquo, family = binomial(), data = ChileYN)

Deviance Residuals:

Min 1Q Median 3Q Max

-3.2095 -0.2830 -0.1840 0.1889 2.8789

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.193759 0.270708 -0.716 0.4741

age 0.011322 0.006826 1.659 0.0972 .

statusquo 3.174487 0.143921 22.057 <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: 2360.29 on 1702 degrees of freedom

Residual deviance: 734.52 on 1700 degrees of freedom

AIC: 740.52

Number of Fisher Scoring iterations: 6

> anova(glmOut, test="Chisq")

Analysis of Deviance Table

Model: binomial, link: logit

Response: vote

Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Pr(>Chi)

NULL 1702 2360.29

age 1 34.2 1701 2326.09 4.964e-09 \*\*\*

statusquo 1 1591.6 1700 734.52 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> exp(coef(glmOut))

(Intercept) age statusquo

0.8238564 1.0113863 23.9145451

> PseudoR2(glmOut)

McFadden Adj.McFadden Cox.Snell Nagelkerke McKelvey.Zavoina Effron Count Adj.Count

0.6888013 0.6854119 0.6150544 0.8201631 0.7855565 0.7553412 0.9230769 0.8433014

AIC Corrected.AIC

740.5206862 740.5348122

> exp(confint((glmOut)))

Waiting for profiling to be done...

2.5 % 97.5 %

(Intercept) 0.4847068 1.402937

age 0.9979335 1.025033

statusquo 18.2483505 32.107663

> bayesLogitOut <- MCMClogit(formula = vote ~ age + statusquo, data = ChileYN)

> summary(bayesLogitOut) # Summarize the results

Iterations = 1001:11000

Thinning interval = 1

Number of chains = 1

Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,

plus standard error of the mean:

Mean SD Naive SE Time-series SE

(Intercept) -0.18272 0.272640 2.726e-03 0.008938

age 0.01123 0.006817 6.817e-05 0.000223

statusquo 3.19061 0.145853 1.459e-03 0.004993

2. Quantiles for each variable:

2.5% 25% 50% 75% 97.5%

(Intercept) -0.742761 -0.365241 -0.17552 -0.0003872 0.34439

age -0.002005 0.006733 0.01121 0.0157683 0.02499

statusquo 2.914442 3.087259 3.18546 3.2847388 3.48698

For both the Bayesian and GLM, it is clear to see improvements from the age and income model run earlier. In this model, both age and status quo are significant predictors. Mostly importantly, for model to model comparison, the AIC in the model developed in the chapter (age+income) the AIC value was 2332 (Stanton, 224). In this model, the AIC has drastically reduced to 740 signifying model improvement. Status quo has a large exponential (regular odds) coefficient at ~24 (95% confidence ranging from 18-32) where income had a regular odds ratio of <1. The Bayesian analysis strengthens these results found in the frequentist method.

We examined data from the 1988 Chilean plebiscite, to see if the age and desire to maintain the status quo of a voter could predict whether an individual would vote in favor of keeping Augusto Pinochet in office. We conducted a logistic analysis using age and status quo to predict votes. We found that both age and status quo (p<.001) predicted votes so that older individuals and those who prefer to maintain the status quo tended to vote for Pinochet. Transforming the log odds for age, (0.01) into plain odds (1.01:1) demonstrates that for every additional year of age a person is 1.01% more likely to vote "Yes". Transforming the log odds for status quo, (3.17) into plain odds (23.9:1) demonstrates that for every additional year of age a person is 23.9% more likely to vote "Yes". The 95% confidence interval for age spans 0.99-1.02 and status quo (18.2-32.1) further providing evidence that age and status quo are significant predictors of voting “Yes”. Finally, the chi-square test validates the claims mentioned previously with age (x2 = 34.2, p<.001) and status quo (1591.6, p<.001) influencing the outcome voting for Pinochet. A confusion matrix demonstrated an error rate of 8% indicating the logistic model was solid at predicting votes.

7)

> plotodds <-function(logodds) {

+ odds <- exp(logodds)

+ hist(odds, breaks =100)

+ abline(v=quantile(odds, 0.025))

+ abline(v=quantile(odds,0.975))

+ }

>

> plotodds(bayesLogitOut[,3])

