Victoria Haley

Homework 9

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
```

Interpretation:

The Deviance of Residuals overlaps with 0. The Y coefficient is significantly different from 0. The coefficients on the ‘gear’ and ‘hp’ variables show the strength of prediction of the Y variable (v-shape or straight engine). The ‘gear ‘ coefficient is observed to not be significantly different from 0 as supported by the z-test and the associated p-value. The ‘hp’ coefficient is observed to be significantly different from 0 as supported by the z-test and the associated p-

value. As such, we fail to reject the null hypothesis that the number of forward gears is 0 but do reject the null hypothesis that the amount of horsepower is not 0.

5: Pseudo R2 Nagelkerke = 0.7789526. Given our finding that only horsepower was significant, these results suggest that horsepower has a significant role in accounting for fuel efficiency.

6: GLM:

```
> ChileGLM <- glm(formula = vote ~ age + statusquo, family=binomial(), data=ChileYN)
> summary(ChileGLM)

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
```

Interpretation:

The age coefficient is not significantly different from 0 based on its z-value of 1.7, nor is it statistically significant based on its p-value of $p > 0.05$. The ‘statusquo’ coefficient is significantly different from 0 based on its z-value of 22, and it is statistically significant based on its p-value of $p < 0.01$. As such, we fail to reject the null hypothesis that the log odds of age is not equal to 0 in the population, but we can reject the null hypothesis that the log odds of statusquo is 0 in the population. The AIC is significantly lower in this model (740.5) than in the original (2332) meaning that error reduction was easier in this model

Bayesian analysis:

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

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
```

Interpretation:

The point estimates for the intercept and coefficients are quite similar to the output from the traditional logistic regression. The HDI for the age coefficient is -0.002 to 0.025. This means that the true population value of age is somewhere within that range, however the range overlaps with 0 and supports that the null hypothesis fails to be rejected. The HDI for the 'statusquo' coefficient is 2.91 to 3.49. This means that the true population value of 'statusquo' is somewhere within that range. Since this does not overlap with 0, this supports that we can reject the null hypothesis that the odds of statusquo are 0.

```
7: MCMCLogitOutHist <- function(bayesLogitOut) {
        varLogOdds < as.matrix(bayesLogitOut[,parameter])
         hist(varLogOdds)
         abline(v=quantile(varLogOdds, c(0.025), col='black'))
         abline(v=quantile(varLogOdds, c(0.975), col='black'))
        }
```