```
# solutions for HW 6
# R code
# math 150
# Fall 2010
# Jo Hardin


sepsis <- read.table("sepsis.csv",header=T,sep=",")
attach(sepsis)

###################
####  From previous HW 5:
###################


sepsis <- read.table("sepsis.csv",header=T,sep=",")
attach(sepsis)

sepsis.bt <- sepsis[(race==1)&(treat==1),]
sepsis.bu <- sepsis[(race==1)&(treat==0),]


sepsis.bt.log <- glm(death30d ~ apache, family="binomial", data=sepsis.bt)
sepsis.bu.log <- glm(death30d ~ apache, family="binomial", data=sepsis.bu)


summary(sepsis.bt.log)

# Call:
# glm(formula = fate ~ apache, family = "binomial", data = sepsis.bi)

# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -1.7138  -0.8855  -0.6819   1.0298   1.9347

# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept) -2.43215    0.69934  -3.478 0.000506 ***
# apache       0.12130    0.03736   3.246 0.001169 **
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# (Dispersion parameter for binomial family taken to be 1)

#     Null deviance: 97.804  on 71  degrees of freedom
# Residual deviance: 84.897  on 70  degrees of freedom
# AIC: 88.897

# Number of Fisher Scoring iterations: 4




summary(sepsis.bu.log)

# Call:
# glm(formula = death30d ~ apache, family = "binomial", data = sepsis.bu)

# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -1.9385  -1.1479   0.7693   1.0287   1.3959
```

```
# Coefficients:
#              Estimate Std. Error z value Pr(>|z|)
# (Intercept) -0.80765    0.66584  -1.213   0.2251
# apache       0.06148    0.03516   1.749   0.0803 .
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# (Dispersion parameter for binomial family taken to be 1)

#     Null deviance: 79.298  on 57  degrees of freedom
# Residual deviance: 75.963  on 56  degrees of freedom
# AIC: 79.963

# Number of Fisher Scoring iterations: 4


# 1

null deviance = -2ln L(pi0) = 97.804
resid deviance = -2ln L(pi-hat) = 84.897

Y = -2 ln (L(pi0)/L(pi-hat)) = 97.804 - 84.897 = 12.907
pchisq(12.907,1) = 0.99967
1-pchisq(12.907,1) = 0.0003273556

The above test is for the null hypothesis that the slope of the logistic regression
is zero:

H0: beta1=0
Ha: beta1 != 0

p-value = 0.00327

Because we wouldn't see data like ours very often (p=0.00327) if beta1=0, we don't
believe that beta1=0.

That is, we can reject the null hypothesis and claim that the slope of the
regression model is positive.

That is, for higher APACHE scores, the probability of dying of sepsis goes up.


# 2

The Wald test is already done for us in the coefficient table:

estimate = 0.1213
SE(est) = 0.03736
test stat = (0.1213 - 0) / 0.03736 = 3.246
p-value = 0.001169

Again our test is:

H0: beta=0
Ha: beta != 0

The small p-value says we would be unlikely to see our data if beta really does
equal zero, so we don't

believe that beta is zero.  We reject beta =0 and claim that there is a positive
relationship between
```

APACHE score and probability of dying of sepsis. We're glad that the results of the Wald test agree with

the likelihood ratio test.

```
# 3a
null deviance = -2ln L(pi0) = 79.298
resid deviance = -2ln L(pi-hat) = 75.963

Y = -2 ln (L(pi0)/L(pi-hat)) = 79.298 - 75.963 = 3.335
pchisq(3.335,1) = 0.93218
1-pchisq(12.907,1) = 0.06782
```

The above test is for the null hypothesis that the slope of the logistic regression is zero:

H0: beta1=0
Ha: beta1 != 0

p-value = 0.06782

Our p-value > 0.05 which indicates that we don't have evidence to reject the null hypothesis. We cannot

say that beta1 is significantly different from zero. For this group, we would probably not keep APACHE

score as a significant predictor.

3b

The Wald test is already done for us in the coefficient table:

estimate = 0.06148
SE(est) = 0.03516
test stat = (0.06148 - 0) / 0.03516 = 1.749
p-value = 0.0803

Again our test is:

H0: beta=0
Ha: beta != 0

Again, we can't reject the hypothesis that beta1=0. Note that the Wald and likelihood ratio tests give

different results but similar directionality (that is #1 & 2 agree and #3a and 3b agree.)

Also note that APACHE score is significant for one subset of the sample and not the other. The difference

can have practical implications: if in fact the effect is real (observed again in other studies), we would

need to model the groups differently if trying to use APACHE score as a predictor of probability of death.

4

Notice that in a 2x2 table, you are comparing a binary explanatory variable to a binary response variable.

Logistic regression is perfectly suited to use both variables as binary. In the logistic regression model,

the beta1 variable will assess whether or not the probability of success is different for the two levels of

the explanatory variable. After fitting the logistic regression, a test (either Wald or LRT) of beta1=0

will be the same as a test of equal odds.