Homework #3

1. (14.20) Refer to Flu shots Problem 14.14.

a. Obtain joint confidence intervals for the age odds ratio $exp(30\beta_1)$ for male clients whose ages differ by 30 years and for the health awareness index odds ratio $exp(25\beta_2)$ for male clients whose health awareness index differs by 25, with family confidence coefficient of approximately .90. Interpret your intervals.

```
##                         2.5 %       97.5 %
## (Intercept)    4.230181e-93 1.122860e+61
## X1             1.558647e+00 5.744237e+01
## X2             6.036501e-03 3.224406e-01
## as.factor(X3)1 2.995946e-08 2.794310e+19
##                         2.5 %       97.5 %
## (Intercept)    1.051879e-77 7.503626e+50
## X1             1.447515e+00 2.924334e+01
## X2             1.414668e-02 3.893810e-01
## as.factor(X3)1 5.375823e-07 1.604090e+16
```

Under the Bonferroni method with family confidence level of 0.9, the confidence interval for $exp(30\beta_1)$ is (1.559, 57.442). That means, keeping health awareness the same, when the age of male clients differ by 30 years, the odd ratio of recieving flu shot range from 1.559 and 57.442. The 90% confidence interval for $exp(25\beta_2)$ is (0.014, 0.389). That means, keeping age the same, when health awareness index of male clients differs by 25, the odd ratio of recieving flu shot range from 0.014 and 0.389.

b. Use the Wald test to determine whether X_3, client gender, can be dropped from the regression model; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the approximate P-value of the test?

```
##
## Call:
## glm(formula = Y ~ X1 + X2 + as.factor(X3), family = binomial(link = logit),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4037  -0.5637  -0.3352  -0.1542   2.9394
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.17716    2.98242  -0.395  0.69307
## X1              0.07279    0.03038   2.396  0.01658 *
## X2             -0.09899    0.03348  -2.957  0.00311 **
```

```
## as.factor(X3)1  0.43397    0.52179   0.832  0.40558
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 134.94  on 158  degrees of freedom
## Residual deviance: 105.09  on 155  degrees of freedom
## AIC: 113.09
##
## Number of Fisher Scoring iterations: 6
```

- H_0: $\beta_3 = 0$
  H_A: $\beta_3 \neq 0$
- Decision rule: If p-value > 0.05, do not reject the null hypothesis. If p-value < 0.05, reject the null hypothesis.
- Conclusion: From the output, p-value for β_3 is $0.40558 > 0.05$. So we fail to reject the null hypothesis, and conclude that client gender cannot be saved in the regression model.

c. Use the likelihood ratio test to determine whether X_3, client gender, can be dropped from the regression model; use $\alpha = .05$. State the full and reduced models, decision rule, and conclusion. What is the approximate P-value of the test? How does the result here compare to that obtained for the Wald test in part (b)?

```
##        G2   p-value
## 0.7022111 0.4020417
```

- H_0: $\beta_3 = 0$
  H_A: $\beta_3 \neq 0$
- Decision rule: If p-value > 0.05, do not reject the null hypothesis. If p-value < 0.05, reject the null hypothesis.
- Conclusion: From the output, p-value for β_3 is $0.4020417 > 0.05$. So we fail to reject the null hypothesis, and conclude that client gender can be dropped from the regression model.

The result of the likelihood ratio test is essentially the same as the result from the Wald test.

d. Use the likelihood ratio test to delermine whether the following three second-order terms, the square of age, the square of health awareness index, and the two-factor interaction effect betwecn age and health awareness index, should be added simultaneously to the regression model containing age and health awareness index first-order terms; use $\alpha = .05$ State the alternatives. Full and reduced models, decision rule, and conclusion. What is the approximate P-value of the test?

```
##        G2   p-value
## 1.5339541 0.6744562
```

- Full model:

  $$log\frac{\pi}{1-\pi} = \beta_0+\beta_1\cdot(age)+\beta_2\cdot(health)+\beta_3\cdot(age)^2+\beta_4\cdot(health)^2+\beta_5\cdot(age)\cdot(health)$$

  Reduced model:

  $$log\frac{\pi}{1-\pi} = \beta_0 + \beta_1 \cdot (age) + \beta_2 \cdot (health)$$

- H_0: $\beta_3 = \beta_4 = \beta_5 = 0$
  H_A: ALOI
- Decision rule: If p-value > 0.05, do not reject the null hypothesis. If p-value < 0.05, reject the null hypothesis.
- Conclusion: From the output, p-value is $0.6744562 > 0.05$. So we fail to reject the null hypothesis, and conclude that the three second-order terms can not be added simultaneously to the regression model.

2. (14.22) Refer to Flu shots Problem 14.14 where the pool of predictors consists of all first-order terms and all second-order terms in age and health awareness index.

c. Find the best model according to the AIC_p criterion.

```
full <- glm(Y~X1+X2+I(X1^2)+I(X2^2)+X1:X2,
            family = binomial(link=logit), data = df)
step(full,direction="both")

## Start:  AIC=116.26
## Y ~ X1 + X2 + I(X1^2) + I(X2^2) + X1:X2
##
##           Df Deviance    AIC
## - I(X1^2)  1   104.31 114.31
## - X1:X2    1   104.71 114.71
## - I(X2^2)  1   105.71 115.71
## <none>         104.26 116.26
##
## Step:  AIC=114.31
## Y ~ X1 + X2 + I(X2^2) + X1:X2
```

```
##
##            Df Deviance    AIC
## - X1:X2    1   104.84 112.84
## - I(X2^2)  1   105.75 113.75
## <none>         104.31 114.31
## + I(X1^2)  1   104.26 116.26
##
## Step:  AIC=112.84
## Y ~ X1 + X2 + I(X2^2)
##
##            Df Deviance    AIC
## - I(X2^2)  1   105.80 111.80
## - X2       1   106.73 112.73
## <none>         104.84 112.84
## + X1:X2    1   104.31 114.31
## + I(X1^2)  1   104.71 114.71
## - X1       1   112.64 118.64
##
## Step:  AIC=111.8
## Y ~ X1 + X2
##
##            Df Deviance    AIC
## <none>         105.80 111.80
## + I(X2^2)  1   104.84 112.84
## + I(X1^2)  1   105.73 113.73
## + X1:X2    1   105.75 113.75
## - X1       1   113.20 117.20
## - X2       1   116.27 120.27
##
## Call:  glm(formula = Y ~ X1 + X2, family = binomial(link = logit), data = df)
##
## Coefficients:
## (Intercept)           X1           X2
##    -1.45778      0.07787     -0.09547
##
## Degrees of Freedom: 158 Total (i.e. Null);  156 Residual
## Null Deviance:     134.9
## Residual Deviance: 105.8  AIC: 111.8
```

The best model is:

$$log\frac{\pi}{1-\pi} = \beta_0 + \beta_1 \cdot (age) + \beta_2 \cdot (health)$$

d. Find the best model according to the SBC_p criterion.

```
n <- nrow(df)
step(full,direction="both", k=log(n))

## Start:  AIC=134.67
## Y ~ X1 + X2 + I(X1^2) + I(X2^2) + X1:X2
##
##             Df Deviance    AIC
## - I(X1^2)  1   104.31 129.66
## - X1:X2    1   104.71 130.05
## - I(X2^2)  1   105.71 131.05
## <none>         104.26 134.68
##
## Step:  AIC=129.65
## Y ~ X1 + X2 + I(X2^2) + X1:X2
##
##             Df Deviance    AIC
## - X1:X2    1   104.84 125.12
## - I(X2^2)  1   105.75 126.03
## <none>         104.31 129.66
## + I(X1^2)  1   104.26 134.68
##
## Step:  AIC=125.12
## Y ~ X1 + X2 + I(X2^2)
##
##             Df Deviance    AIC
## - I(X2^2)  1   105.80 121.00
## - X2       1   106.73 121.94
## <none>         104.84 125.12
## - X1       1   112.64 127.85
## + X1:X2    1   104.31 129.66
## + I(X1^2)  1   104.71 130.05
##
## Step:  AIC=121
## Y ~ X1 + X2
##
##             Df Deviance    AIC
## <none>         105.80 121.00
## - X1       1   113.20 123.33
## + I(X2^2)  1   104.84 125.12
## + I(X1^2)  1   105.73 126.01
## + X1:X2    1   105.75 126.03
## - X2       1   116.27 126.41
##
## Call:  glm(formula = Y ~ X1 + X2, family = binomial(link = logit))
```

```
##
## Coefficients:
## (Intercept)           X1           X2
##    -1.45778      0.07787     -0.09547
##
## Degrees of Freedom: 158 Total (i.e. Null);  156 Residual
## Null Deviance:     134.9
## Residual Deviance: 105.8  AIC: 111.8
```

The best model is same as that of (c):

$$log\frac{\pi}{1-\pi} = \beta_0 + \beta_1 \cdot (age) + \beta_2 \cdot (health)$$

3. I received <u>45</u> points in HW #1.