Matt Isaac

Biostatistics – Homework 4

Due: 21 March, 2018

1. Is there a gender difference in the flu shot rate?
    a. Two-way-table-based method:

Table of Gender by Flu			
Gender	**Flu**		
Frequency Col Pct	**0**	**1**	**Total**
F	18 62.07	11 52.38	29
M	11 37.93	10 47.62	21
Total	29	21	50

        i. Test: We will use the General Associations test to answer this question. Because the table has two rows and two columns, the three main chi-square tests (GA, COR, and MSD) are all equivalent. Fisher's exact test is not needed because the expected cell counts are all greater than 5.
        ii. P-value: 0.4977

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)				
Statistic	**Alternative Hypothesis**	**DF**	**Value**	**Prob**
1	**Nonzero Correlation**	1	0.4599	0.4977
2	**Row Mean Scores Differ**	1	0.4599	0.4977
3	**General Association**	1	0.4599	0.4977

        i. Conclusion: Since our p-value is non-significant ($p > 0.05$), there is no evidence of a difference in flu-shot rate between men and women.

    b. Logistic Regression:
        i. Logistic Regression: $H_0: \beta_{Gender} = 0$
        ii. P-value: 0.4941

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Gender	1	0.4676	0.4941

iii. Conclusion: We have no evidence that gender has a significant effect on flu shot rate ($p < 0.05$).

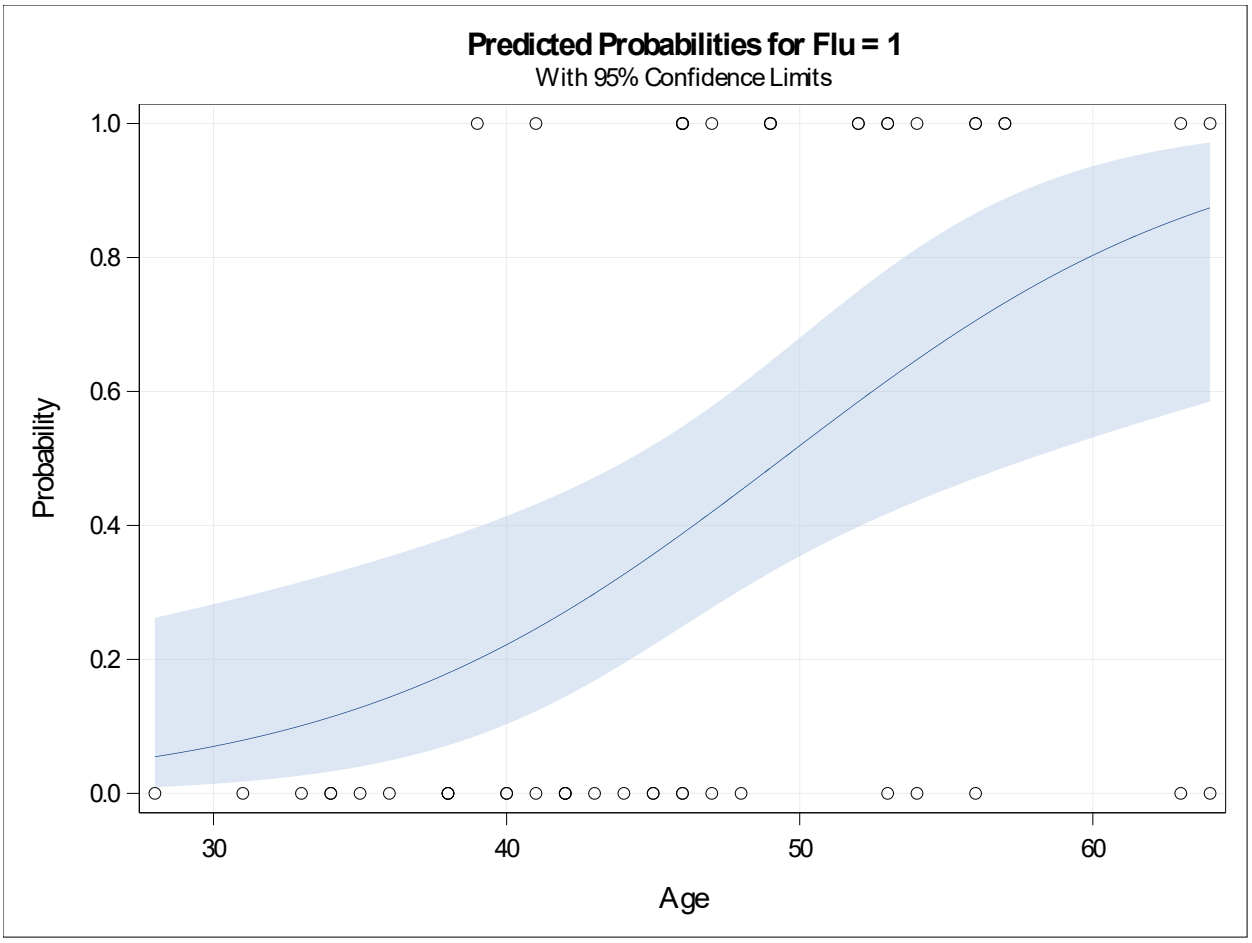
c. Compare results:
   i. The results of the table-based chi-square test and the result of the logistic regression method yielded very similar results; no evidence for a significant difference of flu shot rate between genders by either method. The p-values from our tests were even quite similar (0.4977 and 0.4941).

2. How does age affect the flu shot rate?
   a. Perform an appropriate test of significance:
      i. Test: Carry out through Logistic Regression:
      $H_0: \beta_{\text{Age}} = 0$
      ii. P-value: 0.0030

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.5754	2.1480	9.3707	0.0022
Age	1	0.1330	0.0448	8.8009	0.0030

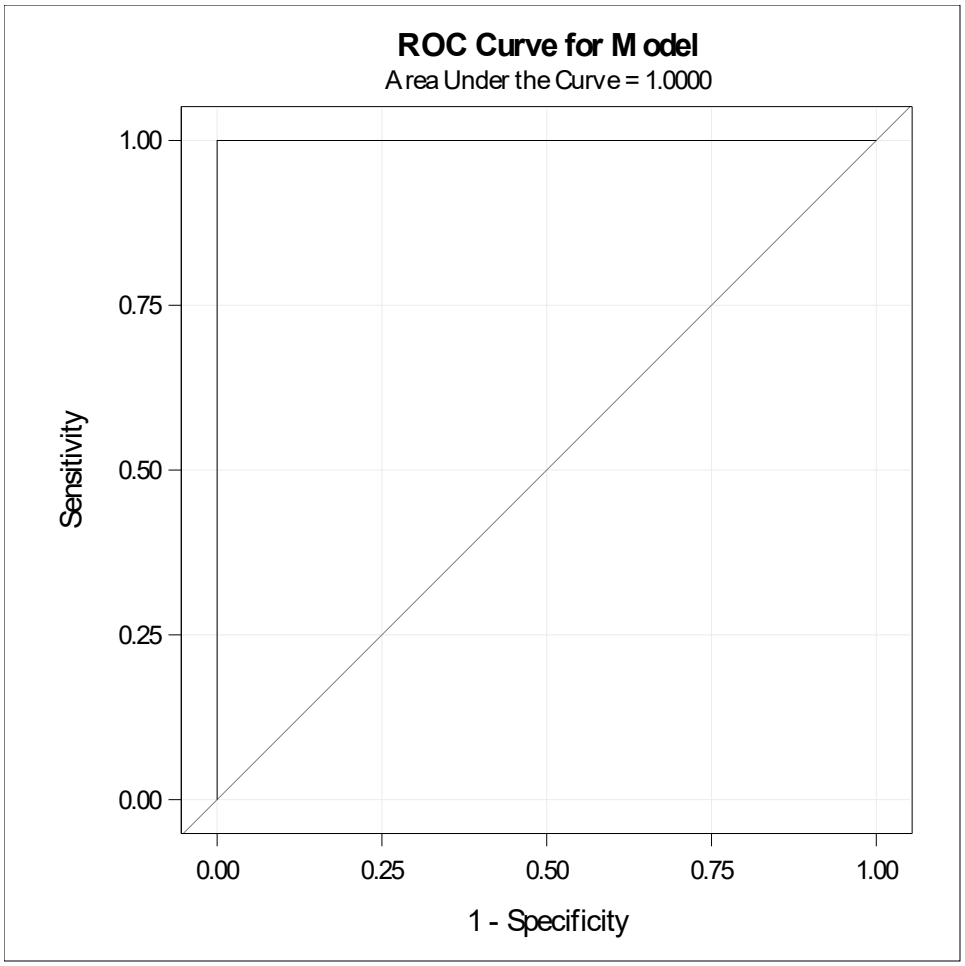
iii. Conclusion: Age does have a significant effect on flu shot rate. The odds ratio is estimated as $\widehat{OR} = e^{0.1330} = 1.14$. After controlling for other factors, the odds of getting a flu shot increases by about 14% for every year increase in age. This is significant ($p = 0.03$).

   b. Dose-response curve:



c. Comment on visualization and test of significance and what they say about the effect of age on flu shot rate.
    i. The visualization and the significant result we obtained suggest that there is a significant increase in the probability of an individual getting a flu shot as age increases. In general, older people are more likely to get a flu shot than their younger counterparts.

3. How do age and health awareness together affect the flu shot rate?
    a. Logistic Regression: test for overdispersion.
        i. Test: Pearson
        ii. P-value: 0.8481
        iii. Conclusion: There is no evidence for overdispersion ($p > 0.05$). We can carry on with the analysis with the assumption that the dispersion parameter is 1.
    b. I have chosen to treat the overdispersion parameter as 1
        i. Age has a significant effect on flu shot rate ($p = 0.0237$). However a person's health awareness does not have a significant effect on flu shot rate ($p = 0.9122$).
    c. ROC curve:
        i. The ROC curve reflects the fact that we got complete separation of points. The AUC = 1, which indicates perfect prediction.



d. Interpret odds ration estimate for age in this model:
    i. We estimate the odds ratio for age as: $\widehat{OR} = e^{0.1509} = 1.16$. This means that, after accounting for health awareness, for every unit increase in age, there is about a 16% increase in the odds of getting a flu shot.

4. (6500)
5. (6500)
6. Accounting for sex differences, is there a difference in tumor response between treatments?
    a. Fit an appropriate model:
        The model we will use will be a multinomial logistic regression model with an ordinal response.
        i. Test: Logistic Regression: $H_0: \beta_{\mathrm{Trt}} = 0$
        ii. P-value: 0.0061
        iii. Conclusion: We reject the null hypothesis ($p < 0.05$). This suggests that, after accounting for sex differences, there is quite a significant difference in tumor response between treatments.
    b. Interpret treatment coefficient in context:

i. The estimated tre5atment coefficient is -0.5807. Holding all else constant, increasing treatment by one unit (going from sequential to alternating treatment) results in a decrease in response of about 0.58. In more general terms, using the alternating treatment as opposed to the sequential treatment leads to an average of a 0.58 shift in tumor response category in a desirable direction (towards tumor gone).

c. Report and test this model's major underlying assumption: The major assumption is the proportional odds assumption. This assumption is met (p = 0.5699).

Appendix: SAS Code

```
/* BioStatistics - Homework 4 */

/* Import data */
FILENAME REFFILE '/home/mattisaac0/BioStatistics/flushot.csv';

PROC IMPORT DATAFILE=REFFILE
        DBMS=CSV
        OUT=WORK.flushot;
        GETNAMES=YES;
RUN;

PROC PRINT DATA = flushot;
run;

/* Exercise 1 - Is there a gender difference in flu shot rate? */

/* (a) two-way-table-based method */
proc freq data = flushot;
        tables Gender * flu / chisq nopercent norow cmh;
        title1 'Testing for gender effect in flu shot rate';
run;

/* (b) logistic regression */
proc logistic data = flushot;
        class Gender;
        model flu(event = '1') = Gender;
        title1 'Logistic Regression - Gender as predictor';
run;

/* Exercise 2 - How does age affect flu shot rate? */

/* (a) Test of significance */
proc logistic data = flushot;
```

```
        class Gender;
        model flu(event = '1') = Age;
        effectplot fit(x = Age);
        title1 'Logistic Regression - Age as predictor';
run;

/* (b) Visualize effect of age - dose response curve */

/* Exercise 3 - How do age and health awareness together affect the flu shot rate? */

/* (a) Perform logistic regression to predict probability of flu shot based on subject's age and health
     awareness. Test for overdispersion in this model by aggregating on both predictors. */
proc logistic data = flushot plots=roc;
        class awareness / param=reference;
        Model flu (event = '1') = age awareness / scale = pearson aggregate=(age awareness) firth;
        Output out=out1 p=phat;
        Title1 'Logistic regression – Age and Awareness';
        Title2 'Check for overdispersion';


/* (b) Account for overdispersion (if present) in logistic regression model. */

/* Exercise 6 - After accounting for sex differences, is there a difference in tumor response between
treatments? */

data lungcancer;
        input trt sex $ response count @@;
        cards;
        0 M 4 28 0 M 3 45 0 M 2 29 0 M 1 26
        0 F 4 4 0 F 3 12 0 F 2 5 0 F 1 2
        1 M 4 41 1 M 3 44 1 M 2 20 1 M 1 20
        1 F 4 12 1 F 3 7 1 F 2 3 1 F 1 1
        ;
run;

/* (a) Fit appropriate model, perform appropriate test of significance. */
proc logistic data = lungcancer;
        class sex/ param=reference;
        model response = trt sex;
        weight count;
        title1 'Ordinal Logistic Regression - Lung Cancer data';
run;
```