Assignment 10 - Logistic Regression. Due December 2, 11:59pm 2018

EPIB607 - Inferential Statistics[a]

[a]Fall 2018, McGill University

This version was compiled on November 21, 2018

In this assignment you will practice logistic regression. Be sure to state the regression model in terms of parameters first. Use regression functions to fit these models. Answers should be given in full sentences (DO NOT just provide the number). All figures should have appropriately labeled axes, titles and captions (if necessary). Units for means and CIs should be provided. All graphs and calculations are to be completed in an R Markdown document. Please submit both the compiled HTML report and the source file (.Rmd) to myCourses by December 2, 2018, 11:59pm. Both HTML and .Rmd files should be saved as 'IDnumber_LastName_FirstName_EPIB607_A10'.

Poisson regression | Rate ratio | Rate difference | Person time | Offset

Template

There is no template for this assignment. You may use the same template from previous assignments. Be sure to include your code at the end of the compiled report.

1. Kidney Stones

The 1986 BMJ article *Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy* by Charig et. al, was a study designed to compare different methods of treating kidney stones in order to establish which was the most cost effective and successful. The procedure, either open surgery, or percutaneous nephrolithotomy (PN, a keyhole surgery procedure), was defined to be successful if stones were eliminated or reduced to less than 2 mm after three months. The study collected cases of kidney stones treated at a particular UK hospital during 1972-1985. The counts of successes for the two surgical procedures, stratified by the size of the kidney stone (smaller than 2cm/at least 2cm in diameter) were:

$<$ 2cm	Unsuccessful	Successful	Total
Open surgery	6	81	87
PN	36	234	270
Total	42	315	357

$\geq$ 2cm	Unsuccessful	Successful	Total
Open surgery	71	192	263
PN	25	55	80
Total	96	247	343

Enter the data into R as frequency records, and complete the following exercises.

a. Fit a logistic regression model for the failure of the surgical procedure given the surgery type and adjusting for the size of the kidney stone.
b. Based on the model output, calculate a 95% confidence interval for the odds ratio of failure of the procedure in open surgery vs. keyhole surgery.
c. Based on the model output, calculate the risks of failure and the expected numbers of failures in all four patient categories. Assess the model fit by comparing the expected failure counts to the observed counts using a χ^2-goodness of fit test. Note that when applying the χ^2-test to a logistic regression model, the test statistic is of the form

$$\sum_{i=1}^{n} \frac{(O_i - E_i)^2}{N_i p_i (1-p_i)} \sim \chi^2_{(n-m)}$$

where n is the number of patient categories, m is the number of estimated parameters and N_i is the total number of patients in the category i. The denominator of the test statistic had to be adjusted because now the risks p_i do not sum to one. In this context, the goodness of fit test is also sometimes referred to as the Hosmer-Lemeshow test.

d. Fit also a saturated logistic model which includes an interaction term between the surgery type and size of the kidney stone. Calculate the expected counts based on this model. Why does it not make sense to test the goodness of fit of a saturated model?
e. Verify that you will get the exact same results as in part (a) by transforming the data into individual level records.
f. Fit also a log-linear model for the risk of failure, given the type of surgery and adjusting for the size of the kidney stone, and based on the results, calculate a 95% confidence interval for the risk ratio of failure of the procedure in open surgery vs. keyhole surgery.