

Homework 4

Aji John

Winter Quarter 2018

Question 1

Perform a regression analysis for evaluating an association between serum creatinine level and 5 year all-cause mortality by comparing the odds of death within 5 years across groups defined by whether the subjects have “high” or “low” creatinine levels, where serum creatinine levels greater than 1.2 are considered to be “high”, (i.e., “high” corresponds to creatinine > 1.2 mg/dl and “low” corresponds to creatinine less than or equal 1.2 mg/dl). In your logistic regression model, use an indicator of death within 5 years as the response, and an indicator of high serum creatinine level as the predictor. (Only provide a formal report of inference when asked to.)

(a)

Is this a saturated regression model? Explain your answer.

Yes, it is a saturated model, as number of parameters equal to number of coefficients.

(b)

Provide an interpretation of the slope and the intercept in your regression model in terms of the response variable (indicator of death within 5 years) and the predictor variable (high creatinine).

Odds when predictor is 0—Found by exponentiation of the intercept from the logistic regression: $\exp(b_0)$ • Odds ratio between groups differing in the value of the predictor by 1 unit—Found by exponentiation of the slope from the logistic regression: $\exp(b_1)$

(c)

From the logistic regression model, what is the estimated odds of dying within 5 years for subjects with low creatinine levels. What is the estimated probability of dying within 5 years from the logistic regression model for subjects with low creatinine levels ?

Odds of dying within 5 years for subjects with low creatinine levels is $e^{-1.8567}=6.402573$

Probability of dying within 5 years for subjects with low creatinine levels is 0.8649118.

- Using $\text{prob} = \text{odds} / (1 + \text{odds})$: $6.402573 / (1 + 6.402573) = 0.8649118$.

(d)

For the subset of subjects in the sample with low serum creatinine, calculate the proportion who died within 5 years. Also calculate the sample odds of dying within 5 years for subjects with low creatinine levels. Compare the sample proportion and sample odds to the corresponding estimates in part 1c from the logistic regression model? Briefly explain any similarities or differences.

Total with Low CRT - 165 Died in 5 years - 44 Proportion who died in 5 years - $44/165 = 0.2666667$

	0	1
--	---	---

Low CRT | 44 | 121

$$\text{Odds} = 1 - (44+121)/121 = 1 - 1.36 = .36$$

(e)

From the logistic regression model, what is the estimated odds of dying within 5 years for subjects with high creatinine levels. What is the estimated probability of dying within 5 years from the logistic regression model for subjects with high creatinine levels ?

(f)

For the subset of subjects in the sample with high creatinine levels, calculate the proportion who died within 5 years. Also calculate the sample odds of dying within 5 years for subjects with high creatinine levels. Compare the sample proportions (or probabilities) and sample odds to the corresponding estimates in part 1e from the logistic regression model? Briefly explain any similarities or differences.

(g)

Give full inference regarding an association between 5 year all-cause mortality and serum creatinine levels from the logistic regression model with indicator of death within 5 years as the response and indicator of high creatinine level as the predictor.

(h)

How would your answers to part b change if you were instead asked to fit a logistic regression model with indicator of death within 5 years as the response variable, but with indicator of low serum creatinine level as the predictor? Would the statistical evidence for an association between 5 year mortality and serum creatinine levels change? Briefly explain.

(i)

How would your answers to part b change if you were instead asked fit a logistic regression model with indicator of surviving at least 5 years as the response variable and indicator of high creatinine level as the predictor? Would the statistical evidence for an association between 5 year mortality and serum creatinine levels change? Briefly explain.

Question 2

In question 1, a prospective association analysis was conducted where we investigated differences in the distribution of death within 5 years across groups defined by serum creatinine level. In this question, you will now conduct a retrospective analysis and fit a logistic regression model for the distribution of serum creatinine across groups defined by vital status at 5 years. In your retrospective logistic regression model, use

an indicator for high serum creatinine level as the response, and indicator of death within 5 years as the predictor. (Only provide a formal report of inference when asked to.)

(a)

Provide an interpretation of the slope and the intercept in your regression model in terms of the response variable (indicator of high creatinine level) and the predictor variable (indicator of death within 5 years).

Our model here is $\log \text{odds} = -1.4214 + 0.8618 * \text{Deathin5i}$

when predictor is 0 (Does not die in 5 years) – Found by exponentiation of the intercept from the logistic regression: $\exp(b_0) = 0.2413759$

Odds ratio between groups differing in the value of the predictor by 1 unit – Found by exponentiation of the slope from the logistic regression: $\exp(b_1) = 2.367418$

(b)

From the logistic regression model, what is the estimated odds of high creatinine level for subjects who die within 5 years? What is the estimated probability of having high serum creatinine for subjects who die within 5 years.

(c)

From the logistic regression model, what is the estimated odds of having a high creatinine level for subjects who survive at least 5 years? What is the estimated probability of having a high serum creatinine for subjects who survive at least 5 years.

(d)

Give full inference regarding an association between 5 year all-cause mortality and serum creatinine levels from the logistic regression model with indicator of high serum creatinine as the response and an indicator of death within 5 years as the predictor.

(e)

Compare the association results in part 2d from the retrospective logistic model to the association results in part 1g from the prospective logistic regression model. Briefly describe any similarities or differences.

Question 3

Perform a regression analysis to evaluate an association between odds of death within 5 years and the continuous measure of serum creatinine levels (i.e., do not use a dichotomized variable for serum creatinine levels in this analysis).

(a)

Provide an interpretation of the slope and the intercept in your logistic regression model.

(b)

Give full inference for an association between 5 year all-cause mortality and serum creatinine levels from your logistic regression model.

eg From logistic regression analysis, we estimate that for two groups that differ by one year in age, the odds of stroke is 3.4% higher in the older group, though this estimate is not statistically significant. A 95% CI suggests that this observation is not unusual if a group that is one year older might have odds of stroke that was anywhere from 0.8% lower or 7.8% higher than the younger group.” A two-sided p value of 0.113 suggests that we can not with high confidence reject the null hypothesis that the odds of stroke are not associated with age.