week7Homework

M. Onimus

10/21/2020

Contents

Read in Data

Question 1 (Logistic Regression)

We are interested in quantifying a possible association between a subject's BMI and whether or not they have been diagnosed with a mental illness in this population. We also wish to control for age, sex, and race (white, black, and other) to account for some basic confounding.

Question/Answer 1a

Given the above model, write out the null and alternative hypotheses for a test of the association between mental illness and BMI.

$log(\frac{\pi}{(1-\pi)}) = \beta_0 + \beta_1 BMI + \beta_2 Age + \beta_3 Sex + \beta_4 Race$

$H_0 : \beta_1 = 0$ $H_1 : \beta_2 \neq 0$

Question/Answer 1b

What is the test statistic for the above test? What is its distribution and associated p-value?

In SAS, the χ^2 = 14.3347 with an associated p-value of 0.0002.

In R, I am provided with a z-value of -3.618 with an associated p-value of 0.000296.

Question/Answer 1c

Write a few sentences describing your findings from this analysis. Be sure to include a discussion about the observed association beyond just the p-value.

When controlling for age, sex, and race, there is a statistical signifcant interaction between BMI and being diagnosed with a mental illness($\chi^2 = 14.3347$ $p - value = 0.0002$). Additionally, the risk of being diagnosed with a mental illness increases by ~6.1% for each $\frac{kg}{m^2}$ increase.

Side note: I am not convinced that this data is actually telling me that BMI and mental illness is related. I think this may be just a correlation in the data and not a true causation event.

Question 2 (Logistic Regression)

A social services program is curious about how common it is for children in the Philadelphia area to be related to the head of the household (or at least, the respondent of the PHMC survey). Therefore, we are interested in the proportion of children in a home that has at least one child.

Question/Answer 2a

How many respondents have at least one child in their home? What is the total number of children living with these subjects? How many of those are related to the respondent? (Hint: In PROC Univariate, the "Sum Observations" cell gives the sum of all observed values.)

There are 290 respondants who have atleast one kid in the home. A total of 612 kids were counted in this study; 550 of those were related to the parents.

Question/Answer 2b

Fit a logistic regression model looking at the number of unrelated children living in a home, as a proportion of the total number of children. Predictors, include just age, sex, and race. Write a few sentences discussing how age, sex, and race are associated with the probability of having an unrelated child in the home, conditional on having at least 1 child present.

Note: my results are different than SAS, not sure why. For the question, I will use the results from SAS

After fitting the model in SAS, both age ($\chi^2 = 53.3$ $p - value =< 0.0001$) and race ($\chi^2 = 22.0$ $p - value =< 0.0001$) are statistically significant when controlling for sex. When probing the results further, race ($\chi^2 = 18.8$ $p - value =< 0.0001$) other is statistically significant when the reference group is white. More specifically, When comparing race other to white, the likelihood of having an unrelated child in the home increased by about 71% (95% CI: 2.558, 11.999).

Question 1 (Survival Analysis)

Note: I have never done survival analysis in R and even though I am really interested, I didn't have time to look up and read how to do it. All the answers below are based on the SAS output.

Question/Answer 1a (Survival Analysis)

Provide a brief discussion of the association between gender and all-cause mortality, including relevant descriptive statistics or figures and using the log-rank test.

The mean years for females were 10.68 while the median was not calculated since it was never reached in females. The mean and median years for males was 11.1 years and 9.47 years, respectively. Comparing just the means, the females average was less by about 0.4 years. When comparing the all-cause mortality by gender, there is a statistically significant difference in the log-rank test ($\chi^2 = 148.5$ $p - value =< 0.0001$) with females surviving longer than males.

Question/Answer 2a (Survival Analysis)

Provide a brief discussion of the association between marital status and all-cause mortality, including relevant descriptive statistics or figures and using the log-rank test.

The mean years for married were 10.50 while the median was not calculated since it was never reached in the married strata. The mean and median years for not married was 9.63 years and 11.60 years, respectively. Comparing just the means, the married strata average was greated by about 0.9 years. When comparing the all-cause mortality by martial status, there is a statistically significant difference in the log-rank test ($\chi^2 = 56.88$ $p - value =< 0.0001$) **with the married strata surviving longer than the not married strata.**

Question/Answer 3a (Survival Analysis)

Provide a brief discussion of the association between self-reported health status and all-cause mortality, including relevant descriptive statistics or figures and using the log-rank test.

People who reported excellent health (mean: 11.12, 25%: 10.86). People who reported very good health (mean: 10.68, 25%: 9.25). People who reported good health (mean: 10.27, 25%: 8.1). People who reported fair health (mean: 9.03, 25%: 5.86, 50%: 10.34). People who reported poor health (mean: 7.30, 25%: 3.15, 50%: 7.36).

When comparing the all-cause mortality by reported health, there is a statistically significant difference in the log-rank test ($\chi^2 = 387.62$ $p - value =< 0.0001$).

Question/Answer 4a (Survival Analysis)

Who really benefits from being married? Provide a discussion for how the effect of marital status on mortality may differ by gender, controlling for age and race.

The first model performed in SAS showed this is no interaction between Married2 and Sex ($\chi^2 = 0.0063$ $p - value = 0.9370$) **and thus was removed from the second model. The study examined the effects of martial status, sex, age, and race on mortality. Using Cox proportional hazards model containing all the parameters, there were significant effects for martial status** ($\chi^2 = 37.34$ $p - value =< 0.0001$), **sex** ($\chi^2 = 166.6540$ $p - value =< 0.0001$), **and age** ($\chi^2 = 899.043$ $p - value =< 0.0001$) **but not race** ($\chi^2 = 2.3322$ $p - value = 0.3116$). **We can estimate that hazards of death decreases by about 24% for married participants (95% CI: 0.696, 0.830) and decreased by about 42% for female participants (95% CI: 0.535, 0.631).**

When comparing the all cause mortality by sex and martial status: - Married females (mean: 11.22, 25%: 10.81) - Married males (mean: 9.74, 25%: 6.68, 50%: 11.84) - Single females (mean: 10.1, 25%: 7.59, 50%: 12.56) - Single males (mean: 7.96, 25%: 4.45, 50%: 8.94)

When comparing the all-cause mortality by martial status and sex, there is a statistically significant difference in the log-rank test ($\chi^2 = 307.5140$ $p - value =< 0.0001$). **These results are expected based on the Cox model.**

To answer the question, males benefit the most from being married with a mean change of ~1.8 years while the female see a smaller mean difference (~1.1 years).