Applied Stat Methods

HOMEWORK 4

In this homework, you will use the **ICU** data set, within the *Stat2Data* package. In class, we have already investigated the relationship between survival status and presence of infection, and the relationship between survival status and age. In this homework, we will look at other variables as predictors of survival.

In all the problems below, our response variable of interest is “Survive”.

1. **Using age group as predictor variable**
   1. EDA: Make a table of survival status by age group, where the values in the table include the *proportion of each age group who survived*.
   2. Does there appear to be a relationship between these two variables? If so, what is the nature of that relationship?
   3. Calculate the odds of survival for each age group.
   4. Calculate and interpret the odds ratio of young people compared to middle-aged people.
   5. Fit a logistic regression model with age group as the explanatory variable. Make sure you tell R that “AgeGroup” is a *categorical* variable.
   6. The intercept of the model is 2.38. Interpret this quantity in context.
   7. For Age Group 2, e^(slope) = 0.33. Interpret this quantity in context.
   8. How does the value in part g (0.33) compare to the OR you calculated in part d?
   9. Is age group a statistically significant predictor of survival in the ICU? Use the drop-in-deviance test to answer this question, and specify the test statistic and p-value.
   10. Calculate a 90% confidence interval for the coefficient on AgeGroup3, and interpret this interval in context.
   11. Is the linearity condition met in this case? Why or why not?
2. **Using pulse as predictor variable**
   1. EDA: Make a boxplot of survival status by pulse.
   2. Does there appear to be a relationship between these two variables? If so, what is the nature of that relationship?
   3. Fit a logistic regression model with pulse as the explanatory variable.
   4. Compute *e*^(slope). What is the interpretation of this quantity?
   5. Is pulse a statistically significant predictor of survival in the ICU? Use the Wald test to answer this question, and specify the test statistic and p-value.
   6. Is the linearity condition met in this case? Why or why not?
3. **Using infection as predictor variable**We’ve already done the EDA for this variable, and calculated the OR of survival for non-infected vs. infected individuals (in Odds & ORs, Part 1).
   1. Fit a logistic regression model with infection as the explanatory variable.
   2. Compute *e*^(slope). What is the interpretation of this quantity?
   3. Is infection a statistically significant predictor of survival in the ICU? Use the Wald test to answer this question, and specify the test statistic and p-value.
   4. Calculate a 95% confidence interval for the slope, and interpret this interval in context.
   5. Is the linearity condition met in this case? Why or why not?
4. **Independence & Randomness**

Consider the conditions of randomness and independence. Does this situation meet those conditions? Discuss. The Examples on pages 476-479 will be very helpful here.

**Article Critique:**

Read the following article:

“Substance Use Among College Athletes:   
A Comparison Based on Sport/Team Affiliation”

You can find this article on the Moodle page in the Folder “Articles” (it’s called “drug use - logistic”).

After reading, briefly **critique** the author’s use of logistic regression.

You’ll want to address questions like…

Is use of the logistic regression model appropriate in this context? What are your thoughts about whether the assumptions of the model are met? Is the sampling method valid, and is the sample representative of the larger population about which he makes conclusions? Are his conclusions valid for the tests he performs? Does he interpret things (like p-values, ORs, and CIs) correctly? Do you feel there are any limitations to this study that he didn’t explicitly discuss?

…But you should feel free to talk about any other issues/concerns that you have about his analysis.