ISyE 6414 Regression Analysis

Homework 7

Fall 2022

For all of the problems in this assignment, please submit the relevant outputs and your R codes (assuming you use R). In addition, unless otherwise indicated, take $\alpha = 0.05$.

You'll see that there are a lot of problems in this assignment. The good news is that the last 5 are for **extra credit!**

For Problems 1–5, refer to the file `homework07data01.csv`, which can be found in the homework section of Canvas. This data set is for a customer satisfication survey for a call center. The data consists of:

- TotalCount — the number of people surveyed, broken down by each of the following categories...
- Time — the amount of time the call took (in min.)
- Resolved — whether or not the customer's issue was resolved (1 = yes)
- Satisfied — whether or not the customer was satisfied with the service

1. Construct a scatter plot of time versus the proportion of satisfied customers with whether or not the issue is resolved (use different colors for resolved and not resolved).

   Construct another scatter plot but now apply the logit function to the proportion.

   Comment on these plots.

2. Construct a logistic regression model using this data set. What are the coefficients? (This is somewhat similar to an example we did in class.)

3. Give an interpretation of the coefficient of the resolved variable with respect to the odds that a customer will be satisfied with the service.

4. Give an interpretation of the coefficient of the time variable with respect to the odds that a customer will be satisfied with the service.

5. Using this model, what are the odds that a customer will be satisfied when their call took 3 minutes and their issue was resolved vs. a customer whose call took 3 minutes and their issue was not resolved. Interpret this.

For Problems 6–15, refer to the file `homework07data02.csv`, which can be found in the homework section of Canvas. This data set consists of shopper data from the trial launch of a new product.

- PriorWeekPurchase — if the customer bought one of the company's products in the previous week
- LastWeekCompPurchase - if the customer bought on of the company's main competitor's products in the previous week
- Age — The age of the customer
- Region — The region in which the store is located
- Count — Number of people surveyed
- Purchase — Number of people who purchased the new product this week
- PurchaseCount — Ignore this for now

6. Construct a logistic regression model based on the probability that the customer will purchase the new product. Use PriorWeekPurchase, LastWeekCompPurchase, Age, and Region as the predictor variables. (Ignore the PurchaseCount column for now — that's for the Poisson model later in the assignment.)

   What are the coefficients?

7. For the model constructed in Problem 6, conduct a hypothesis test for the overall regression.

8. For the model constructed in Problem 6, conduct goodness-of-fit hypothesis tests using (i) the deviance residuals and then (ii) the Pearson residuals. What conclusions would you make using $\alpha = 0.05$? What conclusions would you make using $\alpha = 0.01$?

9. We'll use the same structure as Problem 6, but now ignore the Purchase column in favor of the PurchaseCount column. Our company has introduced a second new product and they believe that this one will be the type that people buy multiple times in a single week, so they're interested in modeling the amount purchased per customer.

   Construct a Poisson regression model using PurchaseCount as the response variable; use PriorWeekPurchase, LastWeekCompPurchase, Age, and Region as the predictor variables, and use Count as the exposure variable. (Ignore the Purchase column for this problem — that was for the Logistic regression model.)

   What are the coefficients of the model?

10. For the model constructed in Problem 9, conduct a hypothesis test for the overall regression.

11. For the model constructed in Problem 9, conduct goodness-of-fit hypothesis tests using (i) the deviance residuals, and then (ii) the Pearson residuals. What conclusions would you make using $\alpha = 0.05$? What conclusions would you make using $\alpha = 0.01$?

12. For the model constructed in Problem **6**, conduct a single hypothesis test to test if the coefficients associated with Age are actually 0.

13. For the model constructed in Problem 9, conduct a single hypothesis test to test if the coefficients associated with Age are actually 0.

14. For the model constructed in Problem **6**, interpret the coefficient of PriorWeekPurchase with respect to the odds that someone will purchase the new product.

15. For the model constructed in Problem 9, interpret the coefficient of PriorWeekPurchase with respect to the rate ratio of the amount of the new item purchased.