AUA Fall 2023

DS 110 Statistics 2

Homework 4

Problem 1 (10 point)

The Poisson distribution may be useful to model events such as

a) Time interval between bus arrivals
b) The price of the plane ticket
c) The time of the first goal during the soccer
d) The number of bus arriving to a bus station between 12:00 and 16:00.

Problem 2 (40 point)

The data diabet.csv has a binary result (outcome, dependent) variable called outcome, which is patient's result of having diabates. There are three predictor variables: Insulin, BMI and Pregnancy. We will treat the variables Insulin and BMI as continuous. The variable pregnancy takes on the values 0 through 11.

a) Construct logistic regression model with dependent variable Outcome and independent variables Insulin, BMI and Pregnancy. Explain coefficients and show which coefficients are significant. (10 points)

b) Find the minimum and maximum values of predictions and discuss how well the model predicts the probability of having diabetes.

c) What is the probability for person with median BMI and Insulin with 0 pregnancy to have diabetes? (10 points)

d) What is odd ratio for patients with maximum BMI and maximum Insulin with 4 pregnancies? (10 points)

Problem 3 (50 point)

In this data armeian_pub.csv, age is the age if the visitor of the pub. Income is a continuous predictor variable and represents visitor's income, and WTS is the maximum willingness to spend at the pub. Frequency is a categorical predictor variable with three levels indicating the frequency of visit to pubs. It is coded as 0 = "Rare", 1 = "Several times a month" and 2 = "Several times a week"

a) Estimate a Poisson regression model for age, including Income, WTI and frequency (as multiple dummy variable) as explanatory variables. Interpret coefficients and the significance of the coefficients (10 points)

b) Obtain predictions from the Poisson regression model. Find the minimum and maximum values and discuss how well the model predicts age. (10 points)

c) Find the expected age of visitor who has 200.000 income, 3000 WTS and is coming to Pub Several times a week. (7 points)

d) Find the expected age of visitor who has 100.000 income, 5000 WTS and is coming to Pub rare. (8 points)

 e) Calculate the probability that at least one of the visitors from c) and d) will be older than 25. (Considering they are independent from each other). (15 points)