Week 3 - AYU - Individual

Son Nguyen

2023-04-06

Type 1: When to use What

## Problem Sets (Apply your Understanding - AYU)

**Problem.** (SRM - Sample Question 11)

Determine which of the following pairs of distribution and link function is the most appropriate to model if a person is hospitalized or not.

1. Normal distribution, identity link function
2. Normal distribution, logit link function
3. Binomial distribution, linear link function
4. Binomial distribution, logit link function
5. It cannot be determined from the information given.

**Problem.** (SRM - Sample Question 14)

From an investigation of the residuals of fitting a linear regression by ordinary least squares it is clear that the spread of the residuals increases as the predicted values increase. Observed values of the dependent variable range from 0 to 100. Determine which of the following statements is/are true with regard to transforming the dependent variable to make the variance of the residuals more constant.

I. Taking the logarithm of one plus the value of the dependent variable may make the variance of the residuals more constant.  
II. A square root transformation may make the variance of the residuals more constant.  
III. A logit transformation may make the variance of the residuals more constant.

1. None
2. I and II only
3. I and III only
4. II and III only
5. The correct answer is not given by (A), (B), (C), or (D).

**Problem.** (SRM - Sample Question 20)

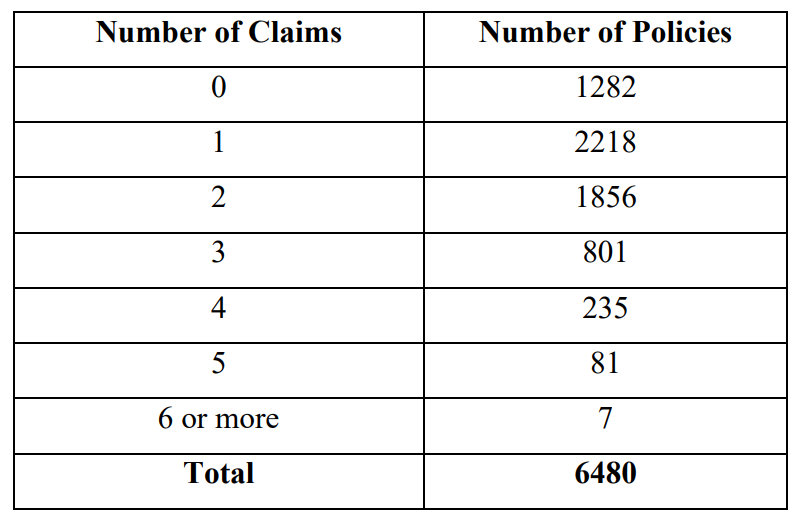
An analyst is modeling the probability of a certain phenomenon occurring. The analyst has observed that the simple linear model currently in use results in predicted values less than zero and greater than one.

Determine which of the following is the most appropriate way to address this issue.

1. Limit the data to observations that are expected to result in predicted values between 0 and 1.
2. Consider predicted values below 0 as 0 and values above 1 as 1.
3. Use a logit function to transform the linear model into only predicting values between 0 and 1.
4. Use the canonical link function for the Poisson distribution to transform the linear model into only predicting values between 0 and 1.
5. None of the above.

**Problem.** (SRM - Sample Question 28)

Dental claims experience was collected on 6480 policies. There were a total of 9720 claims on these policies. The following table shows the number of dental policies having varying numbers of claims.

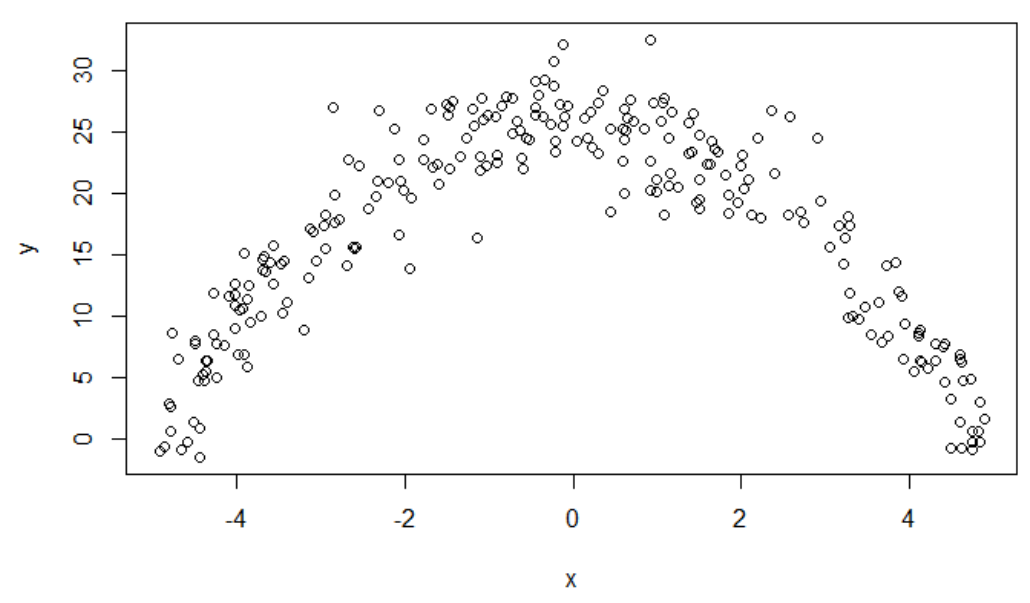


Calculate the chi-squared statistic to test if a Poisson model with no predictors provides an adequate fit to the data.

1. 80
2. 83
3. 86
4. 89
5. 92

**Problem.** (SRM - Sample Question 39)

You are given a dataset with two variables, which is graphed below. You want to predict using . Determine which statement regarding using a generalized linear model (GLM) or a random forest is true.



1. A random forest is appropriate because the dataset contains only quantitative variables.
2. A random forest is appropriate because the data does not follow a straight line.
3. A GLM is not appropriate because the variance of y given x is not constant.
4. A random forest is appropriate because there is a clear relationship between and .
5. A GLM is appropriate because it can accommodate polynomial relationships.

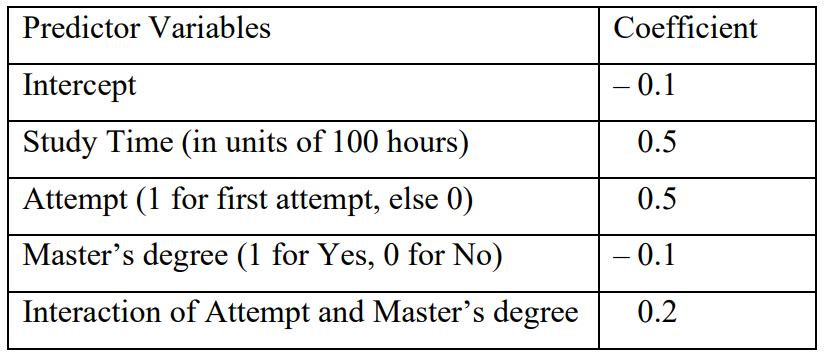
**Problem.** (SRM - Sample Question 42)

Determine which of the following statements is NOT true about the linear probability, logistic, and probit regression models for binary dependent variables.

1. The three major drawbacks of the linear probability model are poor fitted values, heteroscedasticity, and meaningless residual analysis.
2. The logistic and probit regression models aim to circumvent the drawbacks of linear probability models.
3. The logit function is given by
4. The probit function is given by where is the standard normal distribution function.
5. The logit and probit functions are substantially different.

**Problem.** (SRM - Sample Question 45)

The actuarial student committee of a large firm has collected data on exam scores. A generalized linear model where the target is the exam score on a 0-10 scale is constructed using a log link, resulting in the following estimated coefficients



The company is about to offer a job to an applicant who has a Master’s degree and for whom the exam would be a first attempt. It would like to offer half of the study time that will result in an expected exam score of 6.0.

Calculate the amount of study time that the company should offer.

1. 123 hours
2. 126 hours
3. 129 hours
4. 132 hours
5. 135 hours

### Type 1: Calculation

### Type 2: Inference (Testing, prediction intervals…)