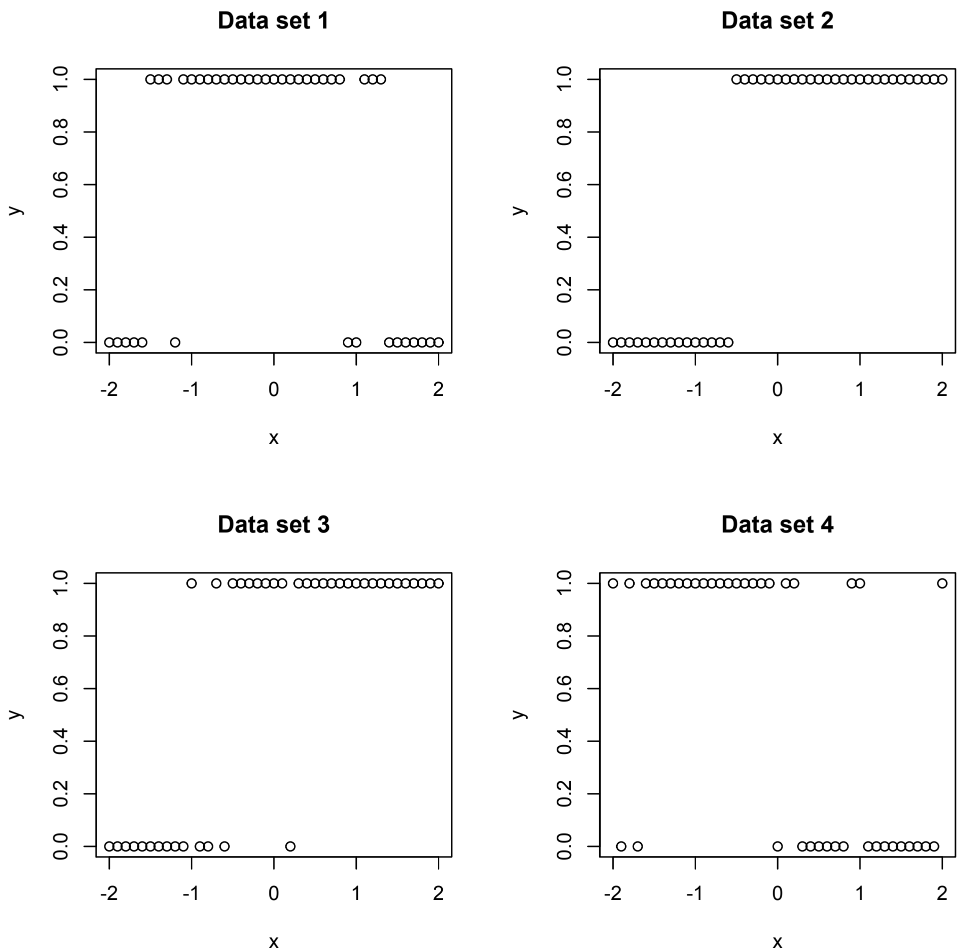
Stat 545 Exam

November 14, 2018

Directions:

- This is a in-class exam. **YOU MUST TURN IN THE EXAM BY 1:50**
- The exam is closed book. You may not use a calculator or a computer or any other electronic device. The exam is designed so that answers requiring numerical calculation can be done accurately enough without a calculator.
- You are allowed one 8.5 by 11 inch page of notes with writing or printing on both sides.
- Use your own paper.
- If you are unable to staple your exam sheets, please fold them to keep them together.
- To get full credit, you must justify your answer.

1. [**50 points**] Below are shown plots of four data sets with a binary response variable Y and a continuous univariate predictor variable x.



Consider the following models for these data sets

Model A `glm(y ~ x, family=binomial(link=probit))`

Model B `glm(y ~ x + x^2, family=binomial(link=logit))`

Model C `glm(y ~ x, family=binomial(link=cauchit))`

Model D None of the above.

Determine which model you think is the best for each data set, and justify your choices.

2. [**50 points**] Suppose we N observations (x_k, y_k), $1 \leq k \leq N$, which we model as i.i.d. realizations of (X, Y) where the discrete random variables X and Y both take values in $\{1, 2, \ldots, I\}$ for a positive integer $I \geq 2$. Assume that both random variables are nominal and the numeric values are codes. We consider 3 different models for the joint distribution of X and Y:

M1 X and Y are independent with the same distribution.

M2 X and Y are independent with possibly different distributions.

M3 X and Y have an arbitrary joint distribution.

Let $\pi_{ij} = P[X = i \,\&\, Y = j]$.

(a) Denoting $\phi_i = \pi_{i+} = \pi_{+j}$ under Model M1, show that the maximum likelihood estimators are given by

$$\hat{\phi}_i \;=\; \frac{1}{2N} \sum_k \left(I[x_k = i] \;+\; I[y_k = i] \right)$$

where $I[x_k = i]$ is 1 if $x_k = i$ and 0 otherwise, and similarly for $I[y_k = i]$.

(b) Describe how to test the null hypothesis H_0 that M1 holds against the alternative hypothesis H_1 that M2 holds. Give a description of how to compute the test statistic and what its approximate null distribution is (i.e., distribution assuming H_0 true). You may write R-code or psuedo-code to show how to compute the test statistic and determine if it is significant.

(c) Repeat (b) with the same null hypothesis but change the alternative hypothesis to H_1 that M3 holds.