

BST 760: Advanced Regression
Breheny

Assignment 4
Due: Thursday, February 21

1. Suppose that

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & -2 \\ 1 & 1 & -1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

For each of the distributions in (a)-(d) below, write \mathbf{W} , the matrix defined on slide 17 of the 2-5 notes. For clarification: do not write formulas or expressions for what \mathbf{W} is; write down the actual numeric values for all of \mathbf{W} 's entries.

- (a) Normal (*i.e.*, linear regression)
 - (b) Poisson
 - (c) Binomial
 - (d) Gamma
2. The course website contains a data set (`education.txt`) representing education expenditure in 1975 for the 50 United States, along with several explanatory variables:
- **Region**
 - **Urban**: Percent of residents living in urban areas
 - **Income**: Per capita personal income
 - **Under18**: Percent of residents under 18 years of age
 - **Expenditure**: Per capita expenditure on public education
- (a) Fit a linear model with **Expenditure** as the outcome variable and all other variables included as additive terms. For the following quantities, provide an estimate, a 95% confidence interval, and a p -value: (i) the effect of increasing the percent of urban population by 10 (ii) the effect of increasing per capita personal income by \$500 (iii) the effect of increasing the percent of the population under age 18 by 2. In addition, report the R^2 of the model.
 - (b) For each region, calculate the mean residual sum of squares. Does education expenditure seem to be equally variable in each region?¹.
 - (c) Use the numbers obtained in (b) to assign weights to each observation according to their region. Refit the model using these weights and repeat part (a).

¹If you are using R for this problem, you can get a visual idea of what is going on with `visreg(fit, "Region")`.