Problem Set 4

Due on May 3

- Submit your answers in a MS Word or PDF file together with the R code to canvas. Use well-complied tables and figures wherever necessary. Besides correctness, the clarity of your word file and R code will also affect grading.

1. Consider the data set hospital_choice.csv with the following variables:

Ducla	dummy for whether patient i goes to UCLA medical center
distance	the distance from patient i's home to UCLA medical center
income	the income of patient i (thousand USD)
old	dummy for whether patient i is older than 75

This is a survey conducted by UCLA medical center to study what kind of patient goes to UCLA medical center for treatment. Use linear, logit, and profit models to estimate the model

$$Ducla \sim income + distance + old.$$

Report the regression results with (McFadden) R-squared. Comment the results.

2. Consider the following data generate process. Let $\beta_1 = 1$, $\beta_2 = 0.5$, $\sigma = 2$, $e_i \sim$ i.i.d.$N(0, \sigma^2)$, $x_i \sim$i.i.d.Gamma(2), and

$$y_i = \beta_1 + \beta_2 x_i + e_i.$$

a. Generate a random sample (e_i, x_i, y_i) with $n = 200$. Use "set.seed(1)" before you generate the data.

b. In an econometric exercise, we observe a x_i and y_i but not e_i. Our parameters of interest is $\theta = (\beta_1, \beta_2, \sigma)$. Suppose we know the parametric family of the condition distribution is normal. From this DGP, we know that $y_i \sim$i.i.d.$N(\beta_1 + \beta_2 x_i, \sigma^2)$, that is

$$f(y_i|x_i, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_1 - \beta_2 x_i)^2}{2\sigma^2}\right).$$

Write down the log-likelihood function and use maximum likelihood to estimate θ.

c. Compute OLS estimate of β_1 and β_2. Compare the results from MLE and OLS.

3. The dataset heating.csv record people choice of heating system. There are five alternatives: gas central (gc), gas room (gr), electric central (ec), electric room (er), heat pump (hp). Here are the explanation of variables

alt	alternatives of heating
ic	installation cost of heating system
oc	operational cost of heating system
income	household income (in unit of $10,000)
agehed	age of household head
rooms	number of rooms of the house
region	categorical variable of house region

a. Which variables indicate individual (decision maker), alternative, and choice? Is the dataset in a long or a wide mode? Use mlogit.data() to turn it into a discrete choice data.

b. Estimate three multinomial logit models using mlogit(). Model (1) only use *ic* and *oc*. Model (2) adds *income*, *agehed,* and *rooms* on top of model (1). Model (3) adds *region* on top of model (2). Note that *region* needs to be transfer into a factor.

Compute the McFadden R2 of each model. Report the results in a table.

c. Use model (3) from part (b). The government try to promote the option of electric central (ec) heating option. There are two plans:

(i) Reduce the installation cost of ec by 10%.

(ii) Reduce the operational cost of ec by 10%.

Fill the following table. Which plan is predicted to be more effective in promoting ec?

	Choice Probability				
	gc	gr	ec	er	hp
Observed					
Predicted by (3)					
Reduce *ic*					
Reduce *oc*					

4. There are three products with the following characteristics and prices

$$X = \begin{array}{c} \begin{matrix} x & y & z & p \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 7 \\ 3 & 1 & 1 & 6 \end{bmatrix} \end{array}.$$

From the dataset product.csv, we observe the sales of these three products from $M = 50$ different geographically separated markets indicated by the *market* variable. Assume that these markets are similar. Although there are only three products ($J = 3$), we can pool information from different market together and estimate the demand. Let the population of consumers be $N = 10000$.

a. Load product.csv into R. Compute the market share of each product at each market.

b. The mean utility is given by

$$\delta_j = \beta_1 x_j + \beta_2 y_j + \beta_3 z_j + \beta_4 p_j,$$

Use the demand estimation approach to estimate $\{\beta_1, \beta_2, \beta_3, \beta_4\}$.

[Hint: given a set of β's, the logit model can help you to generate market shares for three products. This prediction should be the same across all 50 markets.]