MSBA7003 Quantitative Analysis Methods

Assignment 4 (Due October 11 at 23:55; Please submit with the template)

Q1.

Solve the following (super hard) Sudoku problem through mixed integer programming.

A			8				7	
	9			***B***			3	
1					4	9		8
							9	
4				5				
	1		3			6		2
		C						7
	6		2			8		1
		2			3			***D***

Which of the following statement(s) is(are) true.

A) ***A*** = 2.

B) ***B*** = 1.

C) ***C*** = 5.

D) ***D*** = 4.

E) None of the above.

Note: The objective of the Sudoku problem is to fill a 9 × 9 grid with digits so that each column, each row, and each of the nine 3 × 3 subgrids that compose the blocks contain all of the digits from 1 to 9. Now we can use mixed integer programming to solve it.

Decision variables: for each cell, there are 9 possible values, and each value can be chosen (1) or not chosen (0); hence, for the whole 9*9 grid, there are total of 9*9*9 = 729 binary variables. To better formulate the constraints, the decision variables are defined as follows:

$$x_{ijk} = \begin{cases} 1, & if\ the\ value = k\ on\ the\ square\ of\ rows\ i, column\ j \\ 0, & if\ the\ value \neq k\ on\ the\ square\ of\ rows\ i, column\ j \end{cases}$$

Objective: solving of the Sudoku problem depends on the constraints; we can define the objective as constant 0.

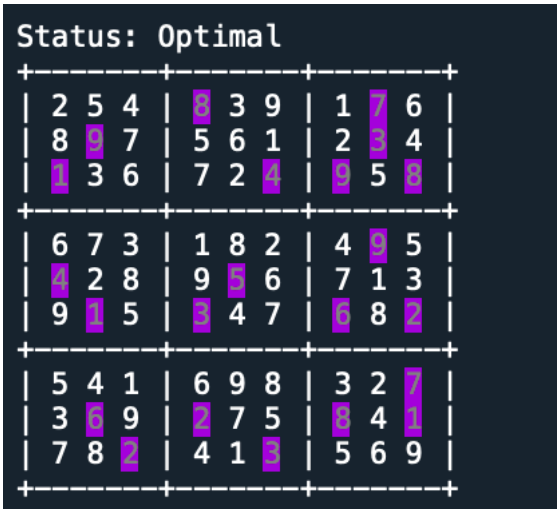
Constrains 1) In each square, only one value can be chosen, i.e., for any fixed square (i,j), $\sum_{k=1}^{9} x_{ijk} = 1$.

2) There are no duplicate numbers in each row, i.e., for any row i=1,2,…0 and a fixed number k, $\sum_{j=1}^{9} x_{ijk} = 1$.

3) There are no duplicate numbers in each column, i.e., for any column j=1,2,…0 and any fixed number k, $\sum_{i=1}^{9} x_{ijk} = 1$.

4) There are no duplicate numbers in each 3*3 subgrids, i.e., denote row i = 3m+a, m = 0,1,2, a = 1,2,3; j = 3n + b, n = 0,1,2, b = 1,2,3; hence, for any fixed number k, m and n, $\sum_{a=1}^{3} \sum_{b=1}^{3} x_{ijk} = 1$.

We use PuLP to solve this problem (please refer to "Q1_Soduku.py" for details.) and obtain the results as follows; hence, A) is True.



Q2.

You are a factory manager and originally the workers are paid a fixed salary according to their skill levels. You want to introduce a productivity-based salary in a hope to increase worker productivity. To begin with, you randomly selected some male and female workers, respectively, according to the numbers given in following Table 1. The selected the workers adopted the new salary scheme in the next month. Their average productivities in the next month are shown in the following Table 2.

Table 1: Worker selection

Gender	Selected	Not Selected	Total Number
Male	15	35	50
Female	65	85	150

Table 2: Average productivity

Male	7 (selected)	6.0 (not selected)

Female	9 (selected)	7.5 (not selected)

Which of the following statement(s) is(are) true?

A) If you implement the new salary scheme for the whole factory, you can expect to increase workers' monthly productivity by 1.25 on average.
B) The naïve estimator of (7*15/80 + 9*65/80 – 6*35/120 – 7.5*85/120) is not biased.
C) Let D = 1 for selected workers and 0 otherwise. Let G = 1 for male workers and 0 otherwise. Suppose we run the regression model: Y = a + b*D + c*G + d*D*G + e, where (a,b,c,d) are parameters and e is the error term. The estimated b should be about 1.5.
D) Among those not selected, their average productivity is expected to be increased by 1.75 if they also adopt the new policy.
E) None of the above.

Note: For choice A), the expect increment is (7-6)*50/(50+150)+(9-7.5)*150/(50+150) =1.375. Choice A) is not True.

For choice B), the average increment of productivities for the male is lower than the average increment of productivities for the female. Since the selection depends on gender, the independence assumption does not hold. Choice B) is not True.

For choice C), we can solve the following equation set: 7=a+b*1+c*1+d*1*1; 6=a+b*0+c*1+d*0*1; 9=a+b*1+c*0+d*1*0; 7.5=a+b*0+c*0+d*0*0. We can get a=7.5, b=1.5, c=-1.5, and d=-0.5. Another way is to regard the value of b as the average increment of female worker productivity. Choice C) is True.

For choice D), the average increment is (7-6)*35/(35+85)+(9-7.5)*85/(35+85)=1.35417. Choice D) is not True.

Q3.

Jenny Wilson Realty is a real estate firm in Alabama. Jenny, the manager, wants to develop a model to determine a suggested listing price based on the size, age, and the condition (either good or excellent) of the house. A sample of historical data includes selling price (Y), the square footage (X1), the age (X2), and the condition (X3 = 1 for being excellent and 0 for being good). Jenny runs a regression of Y against X1, X2, and X3. The estimated model is

$$\hat{Y} = 1.5 + 3X_1 - 0.5X_2 + 0.8X_3.$$

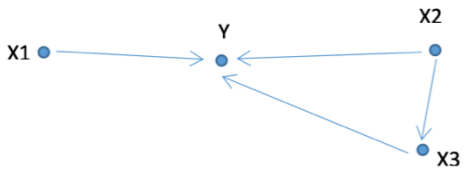
Suppose Jenny would like to find out how much the selling price can be affected by renovating a house and changing the condition from "good" to "excellent". Which of the following statement(s) is(are) true?

A) Renovating a house is expected to increase its selling price by 0.8.
B) Given the same square footage and age but different conditions, two houses' selling prices are expected to differ by 0.8.
C) Given the same square footage and age, the condition of a house is completely independent of any other factors that can affect the selling price.

D) Other variables (such as the location of a house) should be included in the regression in order to estimate how X3 influences Y.

E) None of the above.

Note:



Above is the causal graph among Y, X1, X2, and X3; for choices A) and B), the ages (X2) can also affect the condition (X3). Only given the same square footage and age but different conditions, renovating a house is expected to increase its selling price by 0.8; hence, A) is False, and B) is True.

Since other variables (like the location of a house) determines the quality of the building, which affects both the condition (X3) and selling price of a house (Y), the location of a house should be conditioned on and needs to be added to the regression. Hence, Choice D) is True, and choice C) is False.