MSBA7003 Quantitative Analysis Methods

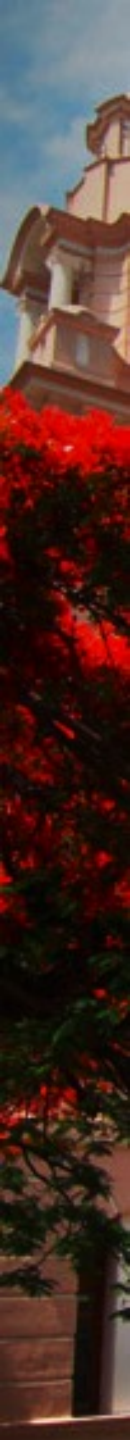
Tutorial 04

2023-2024

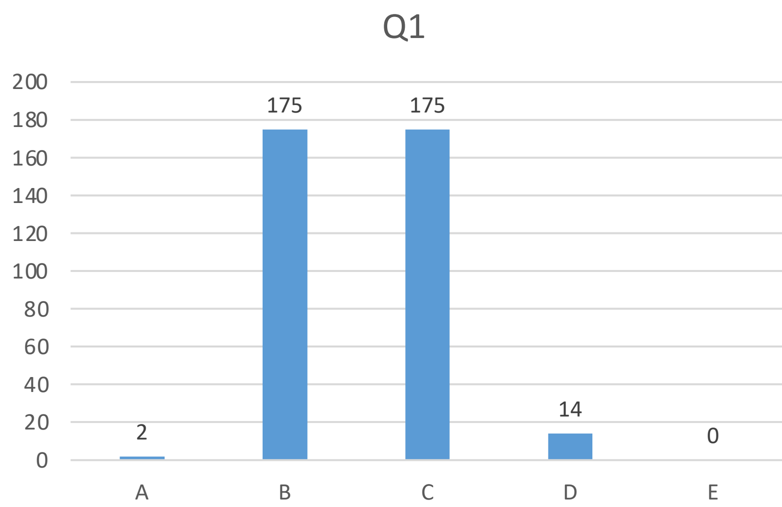
Guoxing HE

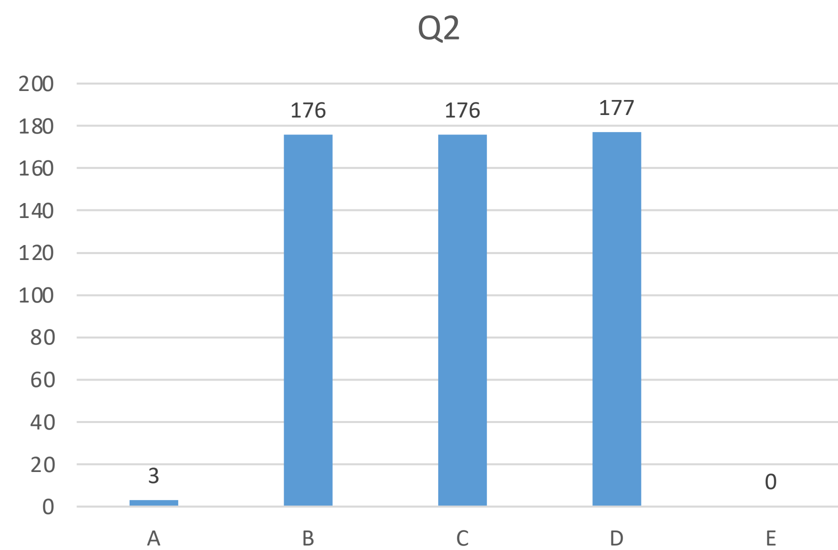
Email: gxhe@connect.hku.hk

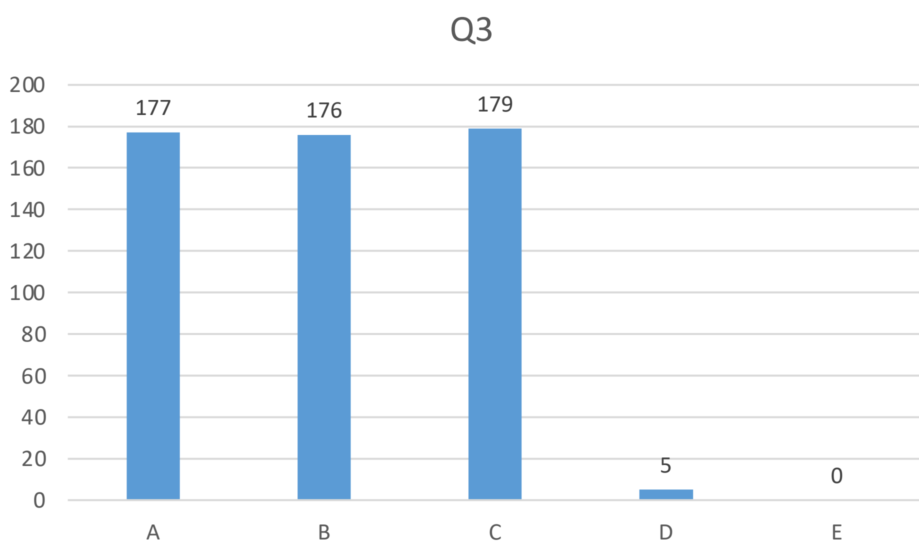
香 港 大 學
THE UNIVERSITY OF HONG KONG



Assignment 3







香 港 大 學
THE UNIVERSITY OF HONG KONG

Agenda

- **Solutions to Assignment 03**
  - Q1 & Q2 – Linear Programming Model
  - Q3 -- Efficiency Evaluation (DEA)
- **Graphical Solution Method for LP**
- **Mixed Integer Programming**
  - Violating Proportionality
- **The Naïve estimator**
- **Causal Inference**
  - Maternal Age and Perinatal Outcomes

Solution to Assignment 03 Q1

Jimmy, an aircraft engineer, is designing the passenger cabin for a new aircraft model. This model will be used by an airline for a specific route. Jimmy needs to decide on the number of seats for both business class and economy class. There must be 4 seats per row in business class; the seat length is 1.2 length units; the seat weight is 2.2 weight units. In economy class, there must be 6 seats per row; the seat length is 1 unit; the seat weight is 1 weight unit. The total cabin length is 32 length units, and the seat weight limit is 200 weight units. Additionally, due to limited demand, business class cannot have more than 6 rows. Jimmy's optimization goal is to maximize the expected profit of the aircraft model for its entire lifecycle on this route. It is known that the profit margin of business class seats is twice that of economy class seats.

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

A) In the optimal design plan, there should be 4 rows for business class and 27 rows for economy class.

B) In the optimal design plan, there should be 5 rows for business class and 26 rows for economy class.

C) In the following problem formulation given by GPT4, constraints 1 – 3 are all wrong.

D) The total weight constraint is not binding.

Solution to Assignment 03 Q1

- let x, y denote the # of rows of business class, and economy class, respectively. We have the following formulation:

$$\max \quad (2*4)x + 6y$$
$$1.2x + y \le 32$$
$$(2.2*4)x + 6y \le 200$$
$$x \le 6$$
$$x, y \ge 0$$

- We use Excel to build and solve the model as follows.

	# of business	# of economy			
Decision Variables	5	26			196
coeff.	8	6			
Constratints	1.2	1	32	<=	32
	8.8	6	200	<=	200
	1		5	<=	6

- Note: sensitivity report is not available if having an integer constraint.

Solution to Assignment 03 Q2

- The Heinlein and Krampf Brokerage firm is instructed by a client to invest \$250,000 among five possible asset classes, as shown in the table below. The return of each security is uncertain and depends on the market scenario. The client listed the following guidelines:
- (1) Investment Grade Bonds should constitute at least 20% of the invested value;

$$x_1 \geq 20\%(x_1 + x_2 + x_3 + x_4 + x_5)$$

- (2) At least 40% of the invested value should be placed in a combination of international equity and S&P 500; $x_4 + x_5 \geq 40\%(x_1 + x_2 + x_3 + x_4 + x_5)$
- (3) No more than 80% of the amount invested in bonds should be placed in hedge funds; $x_3 \leq 80\% x_1$
- (4) No short selling, and it is okay to hold some cash.

$$x_1 + x_2 + x_3 + x_4 + x_5 \leq 250,000$$

Projected Return (%)	in Scenario 1	In Scenario 2
Investment Grade Bonds	1.1	-2.3
Commodities	-5.0	6.8
Hedge Funds	-1.7	4.9
International Equity	-0.3	8.4
S&P 500	-7.5	11.8

Solution to Assignment 03

Projected Return (%)	in Scenario 1	In Scenario 2
Investment Grade Bonds	1.1	-2.3
Commodities	-5.0	6.8
Hedge Funds	-1.7	4.9
International Equity	-0.3	8.4
S&P 500	-7.5	11.8

- The objective is to maximize the worst-case return.
- Define $\boldsymbol{x_i}$ as the amount of money allocated to class i. Denote as $\boldsymbol{r_{ij}}$ as the return rate of class i in scenario j.
- The worst-case payoff is $\mathbf{min}\{\sum \boldsymbol{x_i r_{i1}}, \sum \boldsymbol{x_i r_{i2}}\}$.
- Hence, the LP can be formulated as follows

max $\boldsymbol{t}$

s.t.

$$\boldsymbol{t \leq 1.1x_1 - 5.0x_2 - 1.7x_3 - 0.3x_4 - 7.5x_5}$$
$$\boldsymbol{t \leq -2.3x_1 + 6.8x_2 + 4.9x_3 + 8.4x_4 + 11.8x_5}$$
$$\boldsymbol{x_1 \geq 20\%(x_1 + x_2 + x_3 + x_4 + x_5)}$$
$$\boldsymbol{x_4 + x_5 \geq 40\%(x_1 + x_2 + x_3 + x_4 + x_5)}$$
$$\boldsymbol{x_3 \leq 80\% x_1}$$
$$\boldsymbol{x_1 + x_2 + x_3 + x_4 + x_5 \leq 250{,}000}$$
$$\boldsymbol{x_1, x_2, x_3, x_4, x_5 \geq 0}$$

Solution to Assignment 03

Investment	Investment Grade Bonds	Commodities	Hedge Funds	International Equity	S&P 500			
Scer.1- Projected Return (%)	0.011	-0.05	-0.017	-0.003	-0.075	1350		
Scer.2- Projected Return (%)	-0.023	0.068	0.049	0.084	0.118	4950		
Amount	150000	0	0	100000	0			
Total Return	1350							
Constraints								
	1	1	1	1	1	250000	<=	250000
	1					150000	>=	50000
				1	1	100000	>=	100000
			1			0	<=	120000

A) It is optimal to hold some cash.

B) In the optimal solution, the investment in S&P 500 is zero.

C) In the optimal solution, the worst-case expected return is $1,350.

香 港 大 學
THE UNIVERSITY OF HONG KONG

Solution to Assignment 03

Variable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
B6	Total Return Investment Grade Bonds	1350	0	1	1E+30	1
B4	Amount Investment Grade Bonds	150000	0	0	1E+30	0.009
C4	Amount Commodities	0	-0.061	0	0.061	1E+30
D4	Amount Hedge Funds	0	-0.028	0	0.028	1E+30
E4	Amount International Equity	100000	0	0	0.014	0.0135
F4	Amount S&P 500	0	-0.072	0	0.072	1E+30

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
B6	Total Return Investment Grade Bonds	1350	1	0	3600	1350
B6	Total Return Investment Grade Bonds	1350	0	0	1E+30	3600
G10		150000	0	0	100000	1E+30
G11		100000	-0.014	0	96428.57143	29752.06612
G12		0	0	0	1E+30	120000
G9		250000	0.0054	250000	1E+30	250000

D) In the optimal solution, the return for an additional invested dollar is $0.0054.

Solution to Assignment 03 Q3

- Q3. The Salem Board of Education wants to evaluate the efficiency of the town's four elementary schools. The three outputs of these schools are: (1) average reading score, (2) average mathematics score, and (3) average self-esteem score. The three inputs to these schools are: (1) average educational level of mothers (defined by highest grade completed: 12 = high school graduate, 16 = college graduate, and so on), (2) number of parent visits to school (per child), and (3) teacher-to-student ratio. The relevant information for the four schools is given in the table below.

School	Input 1	Input 2	Input 3	Output 1	Output 1	Output 3
A	14.5	3	0.25	3.5	2.7	3
B	13	2	0.13	3.3	2.5	2.4
C	15.5	4	0.28	3.8	3	3.3
D	16.2	3	0.33	4	3.8	4

Solution to Assignment 03 Q3

- We take school A as an example. For schools B, C, and D, the solving process is similar.
- Assume constant returns to scale

Assuming constant returns to scale						
Evaluating A						
	Input 1	Input 2	Input 3	Output 1	Output 1	Output 3
Weight	0.0689655	0	0	0.251341	0	0.0279693
	TOV		TIV			Efficiency
A	0.9636015	<=	1	fixed as 1		96%
B	0.8965517	<=	0.8965517			100%
C	1.0473946	<=	1.0689655			98%
D	1.1172414	<=	1.1172414			100%

Solution to Assignment 03

- If we assume non-constant returns to scale, we build the following model to check whether school A is efficient or not.

Assuming non-constant returns to scale								
Evaluating A							Weight	
A	14.5	3	0.25	3.5	2.7	3	0	
B	13	2	0.13	3.3	2.5	2.4	0.53125	
C	15.5	4	0.28	3.8	3	3.3	0	
D	16.2	3	0.33	4	3.8	4	0.46875	
Virtual	14.5	2.46875	0.22375	3.628125	3.109375	3.15	Factor k	1
			k-multiple output	3.628125	2.798839	3.109821	1.036607	

- Under the optimal solution, if $K = 1$, school A is efficient; otherwise, it is not efficient.

香 港 大 學
THE UNIVERSITY OF HONG KONG

Graphical Solution Method

Consider the following linear programming problem.

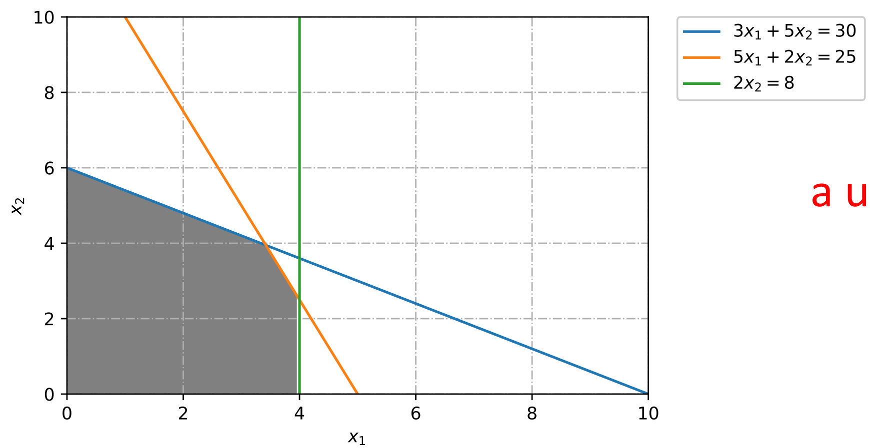
$$\begin{aligned} \max \quad & c'x \\ \text{s.t.} \quad & Ax \leq b, \\ & x \geq 0, \end{aligned}$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $c = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, $A = \begin{bmatrix} 3 & 5 \\ 5 & 2 \\ 2 & 0 \end{bmatrix}$, and $b = \begin{bmatrix} 30 \\ 25 \\ 8 \end{bmatrix}$.

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

- A) The corner points of the feasible region are (x1,x2) = (0,0), (0,3), (4,0), and (4,2.5).
- B) The optimal solution is (x1,x2) = (4,2.5).
- C) There is a redundant constraint.
- D) There are infinitely many optimal solutions.
- E) None of the above.

Graphical Solution Method



a unique optimal solution

- Objective: $\max \quad 3x_1 + x_2$
- Constraint 1: $3x_1 + 5x_2 \leq 30$; Constraint 2: $5x_1 + 2x_2 \leq 25$; Constraint 3: $2x_1 \leq 8$; Constraint 4: $x_1, x_2 \geq 0$.
- Isoprofit line solution method/ Corner point solution method

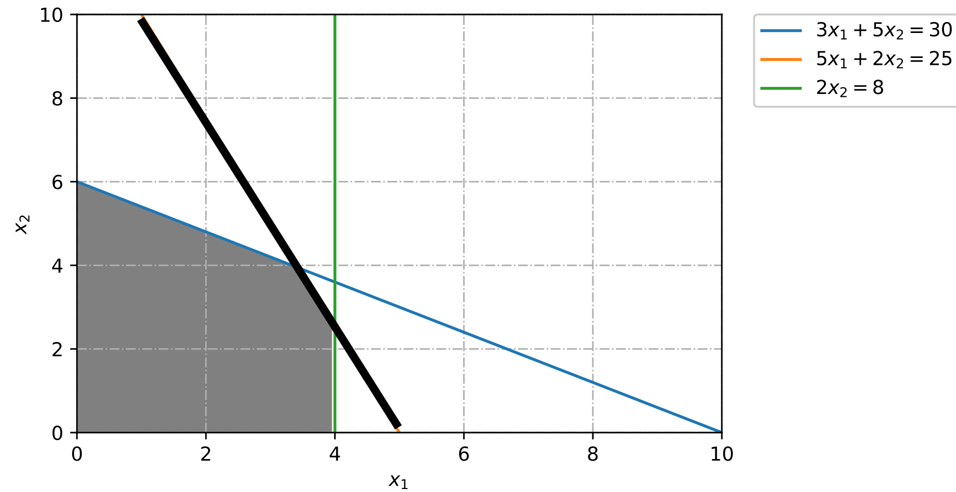
x_1	x_2	$3x_1 + x_2$(max)
0	0	0
4	0	12
4	2.5	14.5
3.42	3.95	14.2
0	6	6

- The feasible region is shaped by all the constraints

香 港 大 學
THE UNIVERSITY OF HONG KONG

Graphical Solution Method

$$\begin{aligned} \max \quad & 6x_1 + 2.4x_2 \\ \text{s.t.} \quad & 3x_1 + 5x_2 \le 30 \\ & 5x_1 + 2x_2 \le 25 \\ & 2x_1 \le 8 \\ & x_1, x_2 \ge 0 \end{aligned}$$

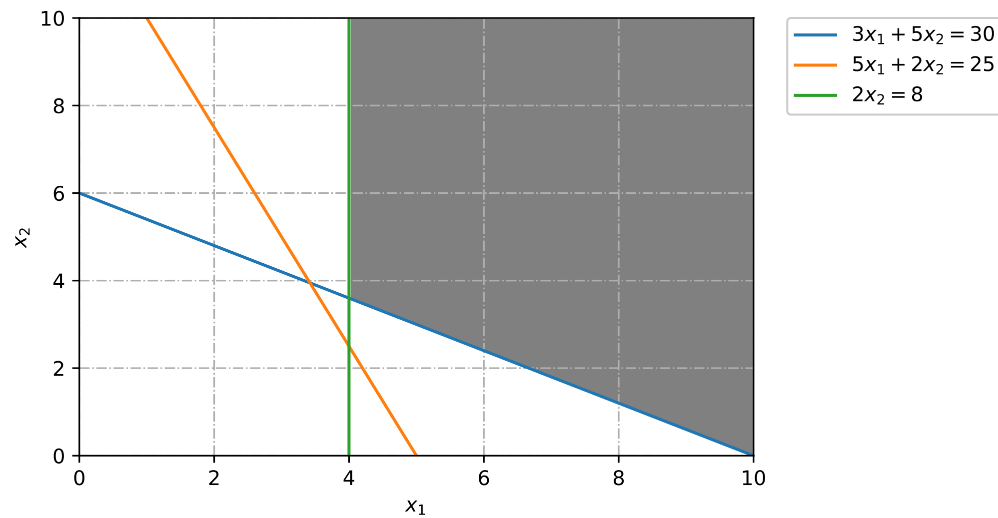


- Extension 1: Objective: $\max \quad 6x_1 + 2.4x_2$

multiple optimal solutions

Graphical Solution Method

$$
\begin{aligned}
\max \quad & 3x_1 + x_2 \\
\text{s.t.} \quad & 3x_1 + 5x_2 \geq 30 \\
& 5x_1 + 2x_2 \geq 25 \\
& 2x_1 \geq 8 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

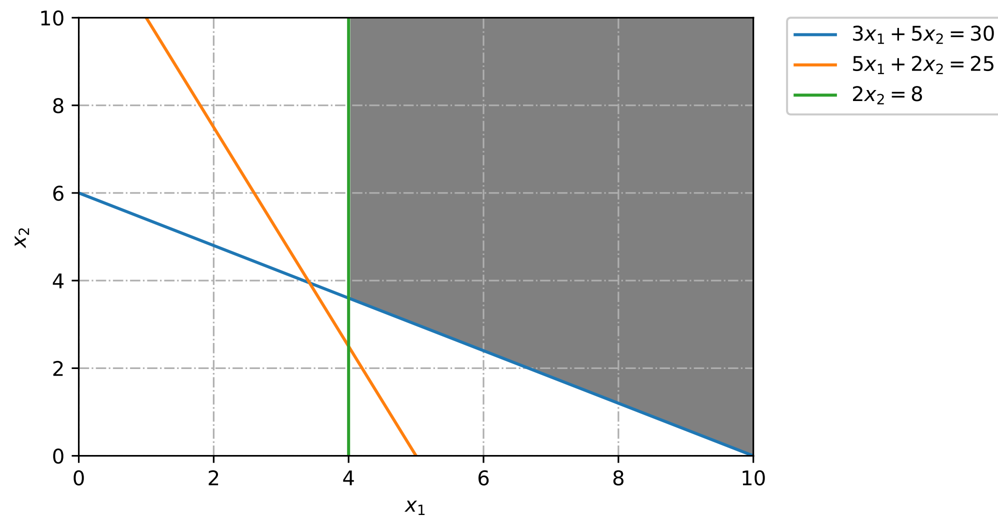


- Extension 2: Constrains: $Ax \geq b$

An unbounded feasible region. Unbounded solution

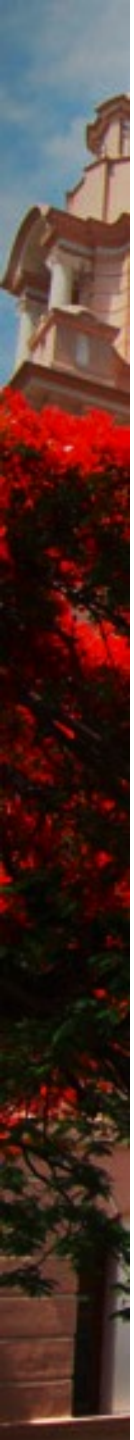
Graphical Solution Method

$$
\begin{aligned}
\min \quad & 3x_1 + x_2 \\
\text{s.t.} \quad & 3x_1 + 5x_2 \geq 30 \\
& 5x_1 + 2x_2 \geq 25 \\
& 2x_1 \geq 8 \\
& x_1, x_2 \geq 0
\end{aligned}
$$



- Extension 3: Objective: $\min \quad 3x_1 + x_2$
  Constrains: $Ax \geq b$

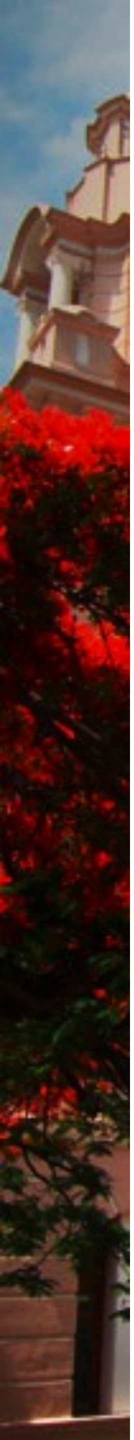
An unbounded region. A bounded solution



Violating Proportionality

- The SUPERSUDS CORPORATION is developing its marketing plans for next year’s new products. For three of these products, the decision has been made to purchase a total of five TV spots for commercials on national television networks. The problem we will focus on is how to allocate the five spots to these three products, with a maximum of three spots (and a minimum of zero) for each product.

Number of TV Spots	Profit		
	Product		
	1	2	3
0	0	0	0
1	1	0	-1
2	3	2	2
3	3	3	4

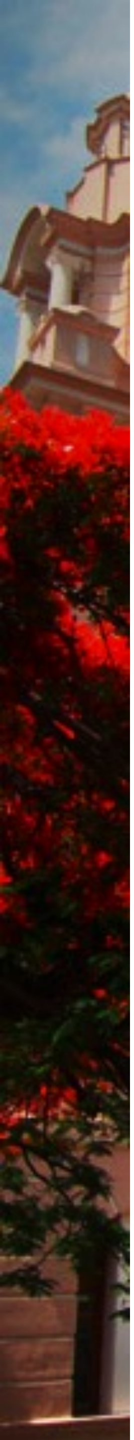


Violating Proportionality

One Formulation with Auxiliary Binary Variables.

- A natural formulation would be to let x_1, x_2, x_3 be the number of TV spots allocated to the respective products.
- However, we cannot write a linear objective function in terms of these integer decision variables. For example, the profit function of product 1, $F_1(x_1)$, is not linear.
- Now we introduce an auxiliary binary variable y_{ij} to replace the original decision variables, where $y_{ij} = \begin{cases} 1 & \text{if } x_i = j \\ 0 & \text{otherwise.} \end{cases}$
- The values of y_{ij} can imply the value of x_i .
- For example, $y_{21} = 0, y_{22} = 0, \text{and } y_{23} = 1$ mean that $x_2 = 3$.
- These definitions are enforced by adding the constraints.

$$y_{i1} + y_{i2} + y_{i3} \leq 1 \text{ for } i = 1, 2, 3.$$



Violating Proportionality

- The resulting linear integer programming model is

$$\text{Maximize} \qquad Z = y_{11} + 3y_{12} + 3y_{13} + 2y_{22} + 3y_{23} - y_{31} + 2y_{32} + 4y_{33},$$

subject to

$$\begin{aligned} y_{11} + y_{12} + y_{13} &\le 1 \\ y_{21} + y_{22} + y_{23} &\le 1 \\ y_{31} + y_{32} + y_{33} &\le 1 \\ y_{11} + 2y_{12} + 3y_{13} + y_{21} + 2y_{22} + 3y_{23} + y_{31} + 2y_{32} + 3y_{33} &= 5 \end{aligned}$$

each y_{ij} is a binary variable.

Violating Proportionality

- The optimal solution is

	y11	y12	y13	y21	y22	y23	y31	y32	y33			
pofits	1	3	3	0	2	3	-1	2	4		total profits	
Decision Variables	0	1	0	0	0	0	0	0	1		7	
Constraints												
	1	1	1							1	<=	1
				1	1	1				0	<=	1
							1	1	1	1	<=	1
	1	2	3	1	2	3	1	2	3	5	=	5

$$
\begin{array}{llllll}
y_{11} = 0, & y_{12} = 1, & y_{13} = 0, & \text{so} & x_1 = 2 \\
y_{21} = 0, & y_{22} = 0, & y_{23} = 0, & \text{so} & x_2 = 0 \\
y_{31} = 0, & y_{32} = 0, & y_{33} = 1, & \text{so} & x_3 = 3.
\end{array}
$$

Violating Proportionality

Another Formulation with Auxiliary Binary Variables.

- We now redefine the above auxiliary binary variables y_{ij} as follows:

$$y_{ij} = \begin{cases} 1 & \text{if } x_i \geq j \\ 0 & \text{otherwise.} \end{cases}$$

- Therefore,

$$\begin{array}{lllll} x_i = 0 & \Rightarrow & y_{i1} = 0, & y_{i2} = 0, & y_{i3} = 0, \\ x_i = 1 & \Rightarrow & y_{i1} = 1, & y_{i2} = 0, & y_{i3} = 0, \\ x_i = 2 & \Rightarrow & y_{i1} = 1, & y_{i2} = 1, & y_{i3} = 0, \\ x_i = 3 & \Rightarrow & y_{i1} = 1, & y_{i2} = 1, & y_{i3} = 1, \end{array}$$

$$\text{so } x_i = y_{i1} + y_{i2} + y_{i3}$$

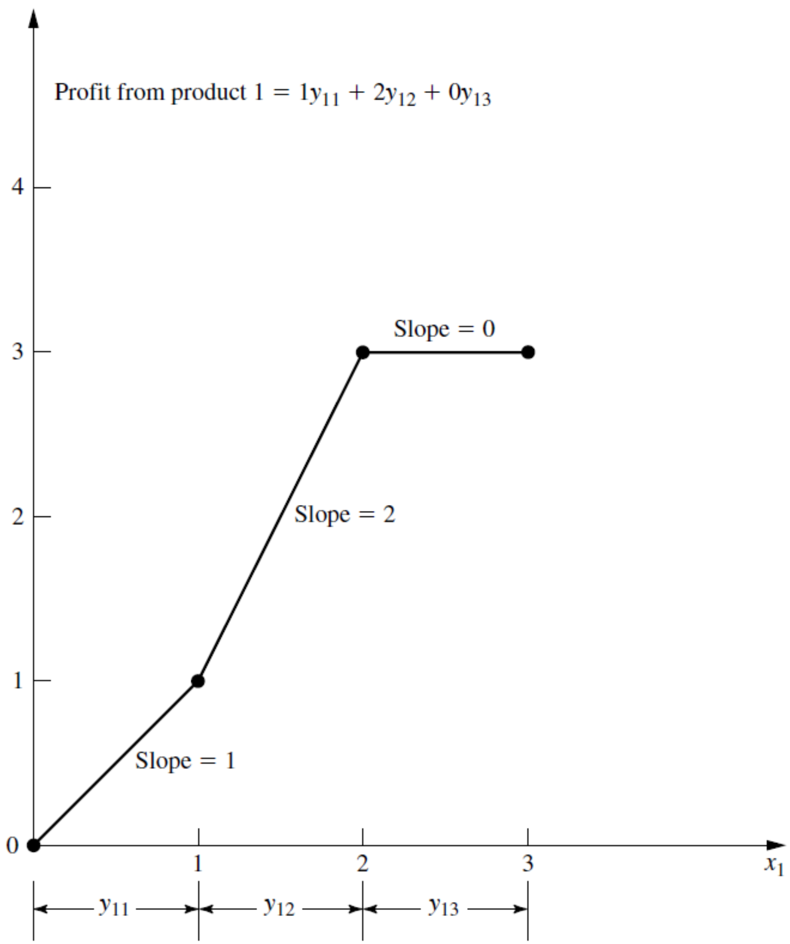
for $i = 1, 2, 3$.

- These definitions are enforced by adding the constraints

$$y_{i2} \leq y_{i1} \quad \text{and} \quad y_{i3} \leq y_{i2}, \quad \text{for } i = 1, 2, 3.$$

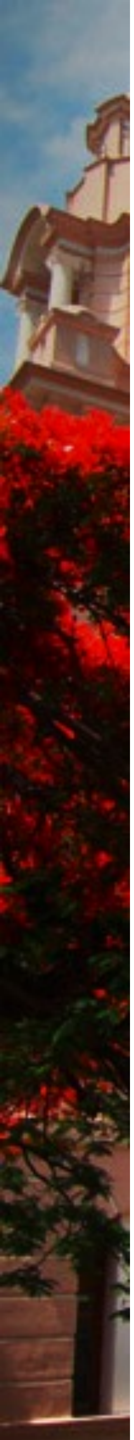
Violating Proportionality

- The new definition of the y_{ij} also changes the objective function.



Number of TV Spots of Product 1	Profit of Product 1
0	0
1	1
2	3
3	3

香 港 大 學
THE UNIVERSITY OF HONG KONG



Violating Proportionality

- The new linear integer programming problem is

$$\text{Maximize} \qquad Z = y_{11} + 2y_{12} + 2y_{22} + y_{23} - y_{31} + 3y_{32} + 2y_{33},$$

subject to

$$\begin{aligned}
&y_{12} - y_{11} \le 0\\
&y_{13} - y_{12} \le 0\\
&y_{22} - y_{21} \le 0\\
&y_{23} - y_{22} \le 0\\
&y_{32} - y_{31} \le 0\\
&y_{33} - y_{32} \le 0\\
&y_{11} + y_{12} + y_{13} + y_{21} + y_{22} + y_{23} + y_{31} + y_{32} + y_{33} = 5
\end{aligned}$$

each y_{ij} is a binary variable.

Violating Proportionality

- Solving this problem gives an optimal solution of

$$
\begin{array}{lllll}
y_{11} = 1, & y_{12} = 1, & y_{13} = 0, & \text{so} & x_1 = 2 \\
y_{21} = 0, & y_{22} = 0, & y_{23} = 0, & \text{so} & x_2 = 0 \\
y_{31} = 1, & y_{32} = 1, & y_{33} = 1, & \text{so} & x_3 = 3.
\end{array}
$$

	y11	y12	y13	y21	y22	y23	y31	y32	y33			
coeff.	1	2	-1	0	2	1	-1	3	2		total profits	
Decision Variables	1	1	0	0	0	0	1	1	1		7	
Constraints												
	1	1	1	1	1	1	1	1	1	5	=	5
		1								1	<=	1
					1					0	<=	0
								1		1	<=	1
			1							0	<=	1
						1				0	<=	0
									1	1	<=	1

Violating Proportionality

- Comparison of two formulations:
  - Decision variables: They have the same number of binary variables (the prime consideration in determining computational effort for IP problems).
  - Constraints: They also have some *special structure:* constraints for *mutually exclusive alternatives* in the first model and constraints for *contingent decisions* in the second.
- The second model does have more functional constraints than the first.
- The first formulation may be better when the problem is on a large scale.

The Naïve estimator

- We would like to estimate the impact of smoking (D) for the general population with a random sample. In the sample, people can be classified along two dimensions. One is smoker versus non-smoker, and the other is high family income versus low family income. The number of people in each class and their average health condition are shown in the following tables.

- What is the Naïve estimator of the impact of smoking on health?

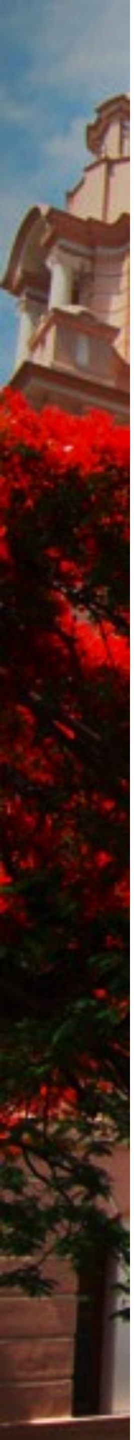
- Why is the Naïve estimator biased?

Number	Smoker	Non-Smoker	Total
High Family Income	3	8	11
Low Family Income	7	3	10
Total	10	11	21

Health Condition	Smoker	Non-Smoker	Average
High Family Income	6	10	8.91
Low Family Income	5	8	5.90
Average	5.30	9.45	

The Naïve estimator

- If we use regression to estimate the impact of smoking based on the given data, what could be the most appropriate model specification (or functional form)?



The Naïve estimator

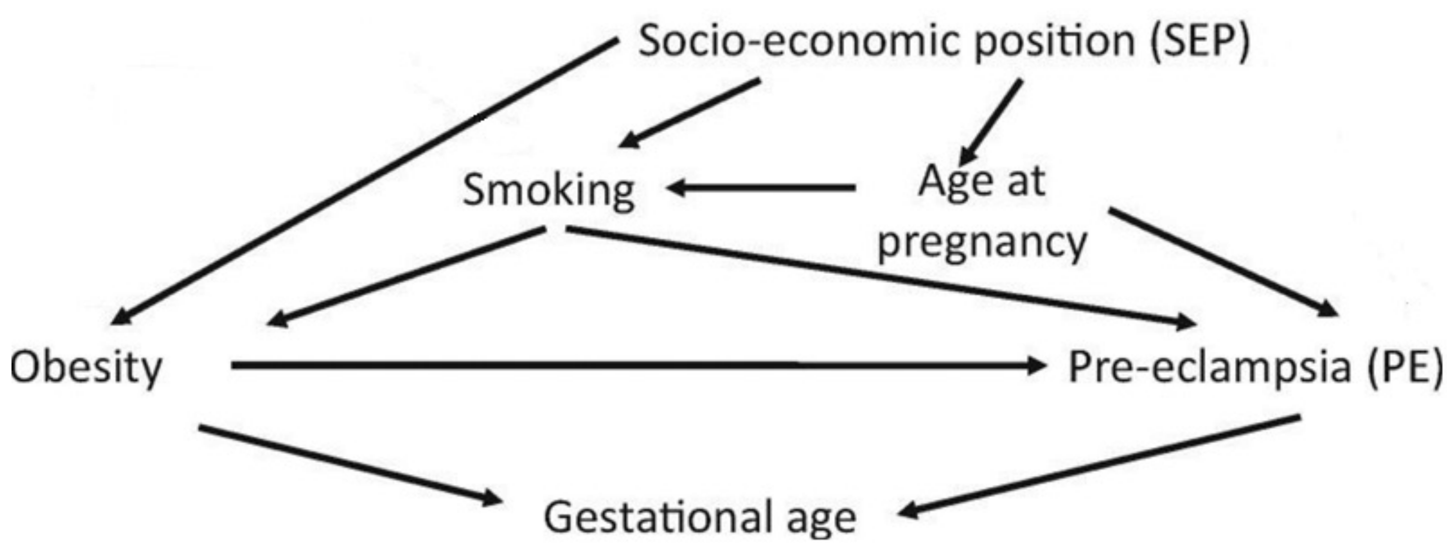
Number	Smoker	Non-Smoker	Total
High Family Income	3	8	11
Low Family Income	7	3	10
Total	10	11	21

Health Condition	Smoker	Non-Smoker	Average
High Family Income	6	10	8.91
Low Family Income	5	8	5.90
Average	5.30	9.45	

If health conditions (Y^0, Y^1) are independent of smoking, conditioning on family income, what is the correct estimate of the impact of smoking on health for the general population?

Maternal Age and Perinatal Outcomes

- We are studying the mechanisms underlying the association between obesity and adverse perinatal (围生期) outcomes (Pre-eclampsia, 妊娠毒血症). We assume that the causal graph is correct [i.e. there are no other variables (nodes) or arrows that should be included].
- How to obtain a valid estimate of the causal effect of obesity on PE.

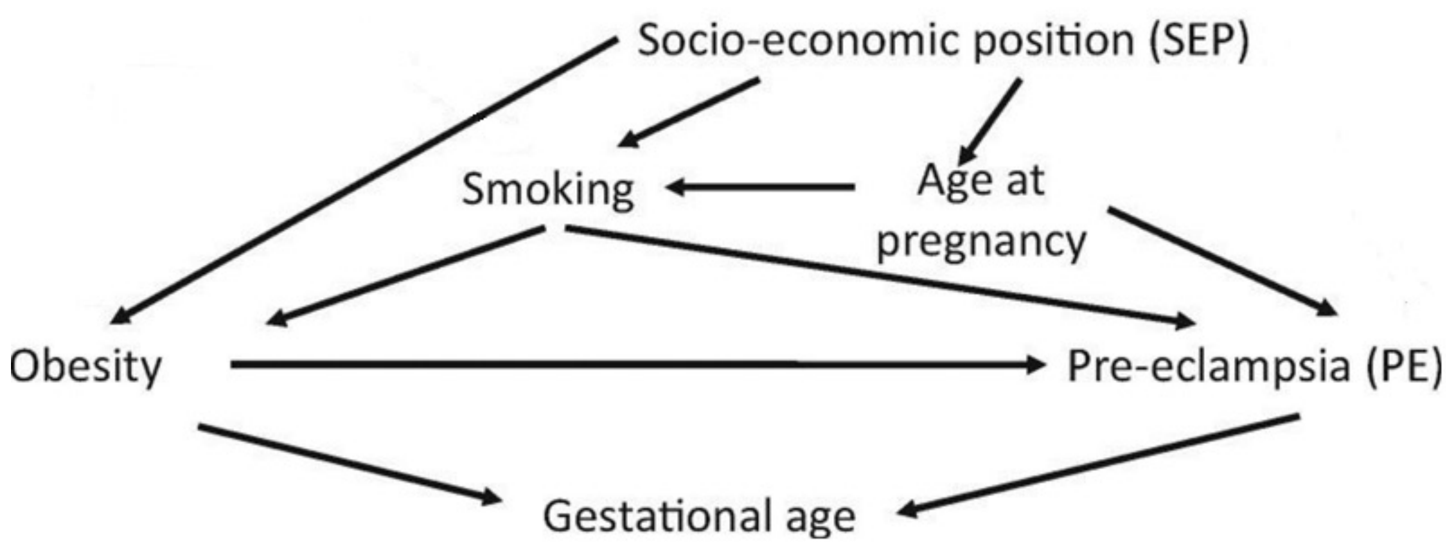


Notes for Back-door-blocking Strategy

- Backdoor paths can transport non-causal associations and must be blocked to avoid spurious associations.
- The backdoor-blocking strategy allows us to identify an appropriate set of covariates for which we should adjust to estimate the causal effect of X on Y
  - Control for a variable = Adjust for a variable = Condition on a variable = using the variable in the regression model
- To estimate the causal effect of X on Y
- Remove all arrows starting from X
  - An unblocked path (without a collider on the way) starting from X transports causes from X to Y
- Identify all unblocked back-door paths
  - unblocked backdoor paths may introduce an association that is not due to the causal effect of X on Y
- Determine whether a set S of covariates is sufficient to block all backdoor paths

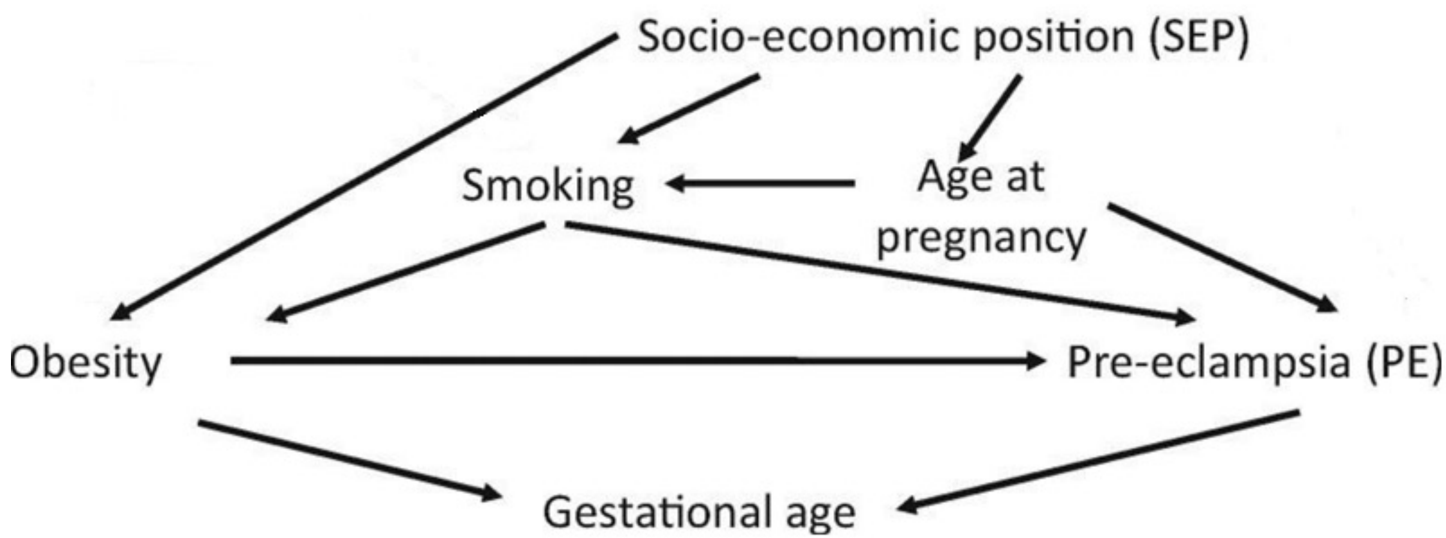
Maternal Age and Perinatal Outcomes

- Unblocked backdoor paths
  - Obesity← Smoking ←SEP → Age at pregnancy → PE
  - Obesity ← SEP → Age at pregnancy → PE;
  - Obesity ← Smoking ← Age at pregnancy → PE;
  - Obesity ← Smoking → PE;
- There is also one blocked path
  - Obesity - SEP- Smoking - Age at pregnancy - PE; this is blocked because age and SEP collide with smoking.



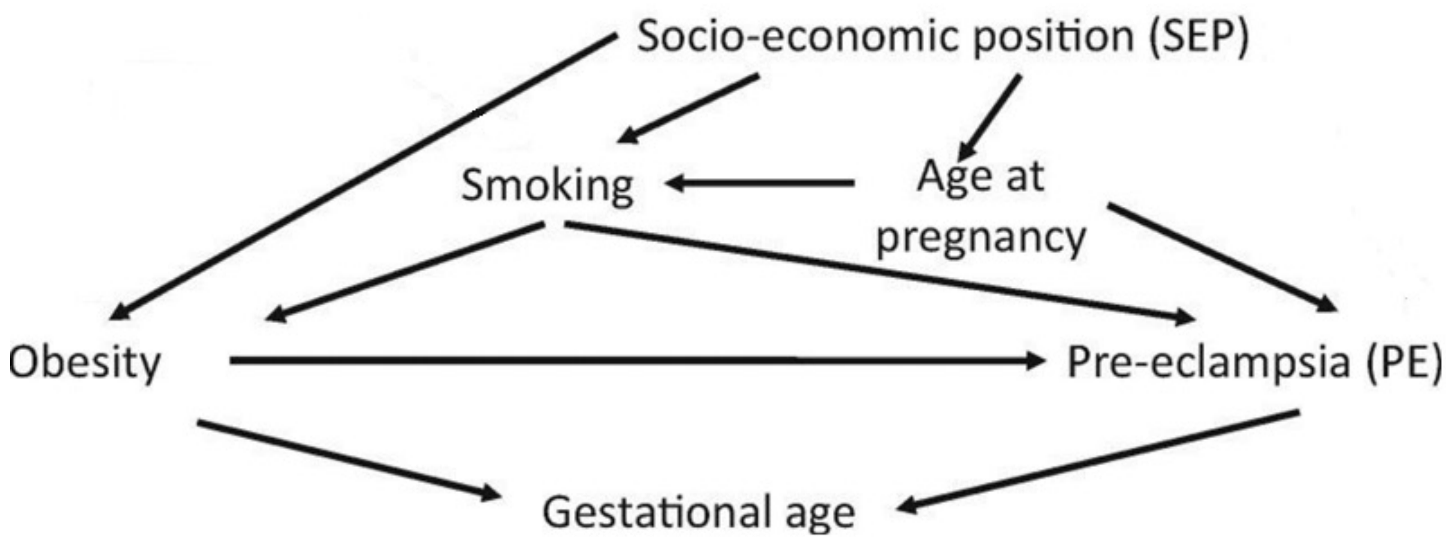
Maternal Age and Perinatal Outcomes

- To block the path Obesity ← Smoking → PE, we must condition "smoking"
- When we do that, a "spurious" association between Pregnancy age and SEP is generated. This new path is different from the above path, i.e., PE-Age at pregnancy-SEP-Obesity.



Maternal Age and Perinatal Outcomes

- However, blocking the original path will block the generated path. To do this, we can condition on either **Age at pregnancy** or **SEP**.
- To conclude, if we assume the causal graph is correct, then conditioning on **Age at pregnancy** (or **SEP**) and **Smoking** will provide a valid causal estimate.



Maternal Age and Perinatal Outcomes

- **Should we condition on gestational?**
- To do so potentially introduces a “spurious” association between Obesity and PE.
- Spurious association would be inverse (or consistent) and so this ‘collider’ bias could produce an effect estimate that is weaker (or stronger) than true positive effect.