Session 4. Advanced LP Models

* Simplex Method

In 1947, George Dantzig invented the first practical algorithm for solving LPs: the simplex method. This essentially revolutionized the use of *linear programming* in practice.

* Data Envelopment Analysis (*DEA*)

DEA is a non-parametric method that has been used to *empirically* measure productive efficiency of producers.

Ex] Is winning everything? How to compare the *efficiencies* of NFL franchises

- Inputs: X_1: Salaries paid to offense players
  X_2: Salaries paid to defense players
- Outputs Y_1: Total number of wins
  Y_2: Offensive yards per attempt (*YPA*)
  Y_3: Defensive yards per attempt against (*YPAA*)

* Goal Programming (*GP*)

GP is an optimization technique to solve problems with multiplicity of objectives, which are generally incommensurable and they often conflict with each other.

Ex] Are you a pet-lover? In a pet shop, rats cost 5 dollars, guppies cost 3 dollars, and crickets cost 10 cents. How many rats, guppies and crickets were sold yesterday?

- Goal #1: 100 animals were sold.
- Goal #2: The total receipts were $100.

A. Simplex Method

* Standard Form of LP

Max $z = c_1x_1 + c_2x_2 + \dots + c_nx_n$
subject to

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$
$$\dots$$
$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

and all $x_i \geq 0$

- *Maximization* in objective function?

  If not, multiply the objective function by -1.

- *Equality* in constraints?

  Add *slack* or *surplus* variables.

- *Non-negative* variables!

  If x is *unrestricted* in sign, then replace it with $(x' - x'')$ where x' and x'' are non-negative variables.

* Matrix Form of LP

Max	$z = \mathbf{c}'\mathbf{x}$	=> Max
subject to	$\mathbf{Ax} = \mathbf{b}$	=> Equality
and	$\mathbf{x} \geq 0$	=> Non-negativity

where $\boldsymbol{c}' = [c_1, c_2, \dots, c_n]$,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix}$$

Ex] Convert the following LP problem to *standard* form.

$$
\begin{array}{lll}
\text{Max} & z = 4x_1 + 3x_2 & \\
\text{subject to} & x_1 + x_2 & \leq 40 \\
 & 2x_1 + x_2 & \geq 60 \\
\text{and} & x_1, x_2 & \geq 0
\end{array}
$$

▪ **Standard form**

where s_1 is a *slack* variable and s_2 is a *surplus* variable.

▪ **Matrix form**

$$
\begin{array}{lll}
\textit{Max} & z = \mathbf{c}'\mathbf{x} & \\
\text{subject to} & \mathbf{Ax} = \mathbf{b} & \Rightarrow \text{Equality} \\
\text{and} & \mathbf{x} \geq 0 & \Rightarrow \text{Non-negativity}
\end{array}
$$

where,

$$
\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & -1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 4 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ s_1 \\ s_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 40 \\ 60 \end{bmatrix}
$$

The optimal solution can be obtained by the simplex method!

* Preview of the Simplex Algorithm

- Consider a linear system **Ax** = **b** of m linear equations with n variables (assume $n \geq m$)
- To find a basic solution to the linear system, we choose a set of n-m variables (non-basic variables) and set each of these variables equal to 0. Then solve for the values of the remaining m variables (basic variables) that satisfy **Ax**=**b**.
- Any basic solution to the linear system in which all variables are non-negative is called a basic feasible solution.
- For any LP, there is a unique extreme point of the LP's feasible region corresponding to each basic feasible solution. Also, there is *at least* one basic feasible solution corresponding to each extreme point of the feasible region.
- For any LP with m constraints, two basic feasible solutions are said to be *adjacent* if their sets of basic variables have m-1 basic variables in common.

Ex] Consider the linear system of 2 equations with 3 variables:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

- If we choose x_1 as a non-basic variable $x_1 = 0$, then the basic variables are $x_2 = 3$ and $x_3 = 2$.
- If we choose x_2 as a non-basic variable $x_2 = 0$, then the basic variables are
- If we choose x_3 as a non-basic variable $x_3 = 0$, then the basic variables are

Ex 1] Primal LP: Consider the LP problem with two variables.

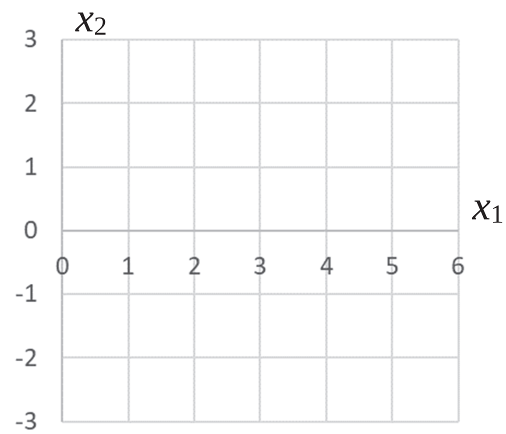
Max $z = 3x_1 + x_2$

subject to

$x_1 + 2x_2 \leq 6$

$x_1 \quad - x_2 \leq 3$

$x_1, x_2 \geq 0$



(a) Convert the LP problem into its matrix form.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}, \ \mathbf{c} = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ s_1 \\ s_2 \end{bmatrix}, \text{ and } \ \mathbf{b} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$$

(b) What is the maximum number of basic solutions?

(c) Evaluate each extreme point.

	x_1	x_2	s_1	s_2	Feasible?	*Max z*
A	0	0	6	3	Yes	0
B		**0**		**0**		
C	4	1	0	0	Yes	**13**
D	0	3	0	6	Yes	3
E	6	0	0	-3	*No*	18
F	0	-3	12	0	*No*	-3

Ex 2] Dual LP: Consider the following LP problem.

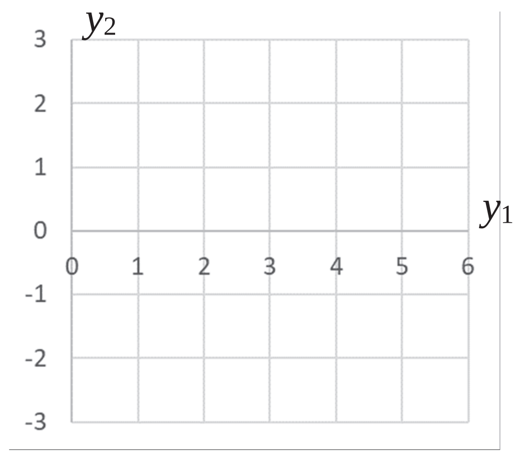
Min $z = 6y_1 + 3y_2$

subject to

$y_1 + y_2 \geq 3$

$2y_1 - y_2 \geq 1$

$y_1, y_2 \geq 0$



(a) Convert the LP problem into its matrix form.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 & 0 \\ 2 & -1 & 0 & -1 \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} -6 \\ -3 \\ 0 \\ 0 \end{bmatrix},\ \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ t_1 \\ t_2 \end{bmatrix},\ \mathbf{c} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

(b) What is the maximum number of basic solutions?

(c) Evaluate each extreme point.

	y_1	y_2	t_1	t_2	Feasible?	*Min z*
A	0	0	-3	-1	*No*	0
B	**0**		**0**			
C	4/3	5/3	0	0	Yes	**13**
D	1/2	0	-7/2	0	*No*	3
E	3	0	0	5	Yes	18
F	0	-1	-4	0	*No*	-3

* Simplex Algorithm

- Step 1. Convert the LP to *standard* form.
- Step 2. Find a basic feasible solution from the standard LP.
- Step 3. Determine whether the current vertex (i.e., basic feasible solution) is optimal.
- Step 4. If not, determine which non-basic variable should become a basic variable and which basic variable should become a non-basic variable in order to find a new basic feasible solution with a better objective function value.
- Step 5. Find the new basic feasible solution with the better objective function value z and go to step 3.

* Efficiency of the Simplex Algorithm

- If an LP has m constraints and n variables, a set of n-m non-basic variables can be chosen in $C(n, m)$ different ways. Thus, an LP can have at most $C(n, m)$ basic solutions.
- Since some basic solutions may not be feasible, an LP can have at most $C(n, m)$ basic *feasible* solutions.
- This means that the simplex algorithm will find the optimal solution after a *finite* number of calculations. (The optimal solution is usually found after examining fewer than $3m$ basic feasible solutions!)

Ex] An LP in standard form with 20 variables and 10 constraints has $C(20, 10)$ = 184,756 basic solutions. The simplex algorithm will find the optimal solution after examining fewer than $3m$ = 30 basic feasible solutions!

More *efficient* algorithms than the simplex method?

Ellipsoid method; Karmarkar's interior point method

B. Data Envelopment Analysis*

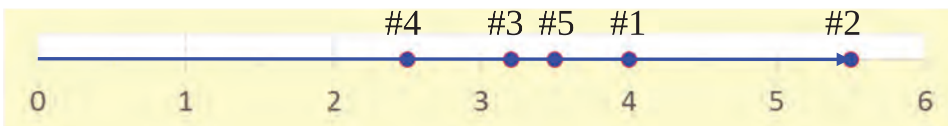
* Introduction

- Data envelopment analysis (*DEA*), occasionally called frontier analysis, is commonly used to evaluate and compare the efficiency of a number of producers.
- A typical statistical approach is characterized as a central tendency approach, and it evaluates producers relative to an average producer. In contrast, *DEA* is an extreme point method and compares each producer with only the "best" producers.
- In the *DEA* literature, a producer is usually referred to as a decision-making unit or *DMU*. Examples of such units are: banks, police stations, hospitals, tax offices, prisons, defense bases, schools, and university departments. (The *DEA* can be applied to non-profit organizations.)
- Several producers can be combined to form a composite producer with composite inputs and composite outputs. Since this composite producer does not necessarily exist, it is sometimes called a virtual producer.
- The heart of the analysis lies in finding the "best" virtual producer for each real producer. If the virtual producer is better than the original producer by either making more output with the same input or making the same output with less input, then the original producer is inefficient.
- The procedure of finding the best virtual producer can be formulated as a linear program. Analyzing the efficiency of n producers is then a set of n linear programming problems.

Ex 1] Single-Input, Single-Output Model

Consider 5 bank branches. For each branch, we have a single output measure (i.e., number of transactions completed) and a single input measure (i.e., number of staff members).

Branch #*i*	Total transactions (output)	Number of staff (input)	Transaction per staff member (ratio)	Efficiency relative to the best
#1	120	30	4.0	**72.73%**
#2	165	30	5.5	100.00%
#3	64	20		
#4	75	30	2.5	45.45%
#5	140	40	3.5	63.64%



- Optimization model for the efficiency of Branch #1:

$Max\ z_1 = \frac{120u_1}{30w_1}$ subject to u_1 and $w_1 \geq 0$,

$\frac{120u_1}{30w_1} \leq 1,\ \frac{165u_1}{30w_1} \leq 1,\ \frac{64u_1}{20w_1} \leq 1,\ \frac{75u_1}{30w_1} \leq 1,\ \frac{140u_1}{40w_1} \leq 1.$

- LP formulation

$Max\ z_1 = 120\ u_1$ subject to u_1 and $w_1 \geq 0$,

$30\ w_1 = 1,\quad 120\ u_1 - 30\ w_1 \leq 0,\quad 165\ u_1 - 30\ w_1 \leq 0,$

$64\ u_1 - 20\ w_1 \leq 0,\quad 75\ u_1 - 30\ w_1 \leq 0,\quad 140\ u_1 - 40\ w_1 \leq 0.$

- Optimal solution: $w_1 = 1/30$ and $u_1 = 1/165$

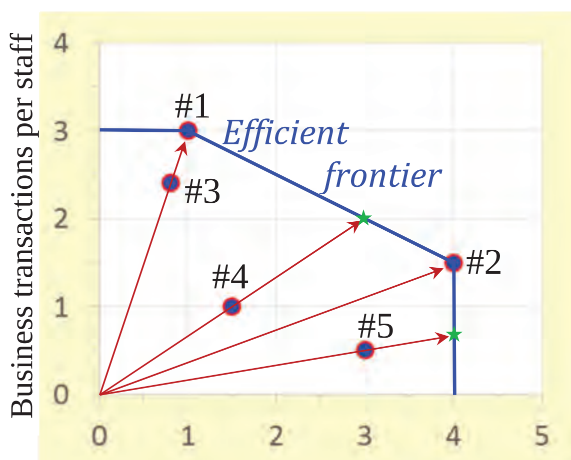
- Efficiency of Branch #1: $z_1 = \frac{120u_1}{30w_1} =$

Ex 2] Single-Input, Multiple-Output Model

Suppose that we have 2 output measures (i.e., number of *personal* transactions completed, and number of *business* transactions completed) and the same single input measure (i.e., number of staff) as before.

Branch #*i*	Output		Input		
	Personal transactions	Business transactions	Number of staff	Personal transactions / staff	Business transactions / staff
#1	30	90	30	1.0	3.0
#2	120	45	30	4.0	1.5
#3	16	48	20	0.8	2.4
#4	45	30	30		
#5	120	20	40	3.0	0.5

▪ **Graphical analysis**



a. Distance from the origin

Branch	Length of vector
#1	3.162
#2	4.272
#3	2.530
#4	
#5	3.041

b. Efficient frontier?

Reference set =

c. Efficiency of each branch

Branch #*i*	#1	#2	#3	#4	#5
Distance from origin to the point	3.162	4.272	2.530	1.803	3.041
Distance from origin to efficient frontier	3.162	4.272	3.162		
Efficiency of each point (i.e., branch #*i*)	100%		80%		75%

- **Efficiency**: Output-input ratio

Branch #i	Output		Input	Output/Input
	Personal transactions	Business transactions	Number of staff	Efficiency
#1	30	90	30	$\frac{30u_i + 90v_i}{30w_i}$
#2	120	45	30	$\frac{120u_i + 45v_i}{30w_i}$
#3	16	48	20	$\frac{16u_i + 48v_i}{20w_i}$
#4	45	30	30	$\frac{45u_i + 30v_i}{30w_i}$
#5	120	20	40	$\frac{120u_i + 20v_i}{40w_i}$

- **Optimization model** for the efficiency of Branch #4:

$Max\ z_4 = \frac{45u_4+30v_4}{30w_4}$ subject to

$\frac{30u_4+90v_4}{30w_4} \le 1,\quad \frac{120u_4+45v_4}{30w_4} \le 1,\quad \frac{16u_4+48v_4}{20w_4} \le 1,$

$\frac{45u_4+30v_4}{30w_4} \le 1,\quad \frac{120u_4+20v_4}{40w_4} \le 1,$ and u_4, v_4, and $w_4 \ge 0$.

- **LP formulation** for the efficiency of Branch #4:

$Max\ z_4 = 45\,u_4 + 30\,v_4$ subject to

$30\,w_4 = 1,$

$30\,u_4 + 90\,v_4 - 30\,w_4 \le 0,\quad 120u_4 + 45v_4 - 30w_4 \le 0,$

$16\,u_4 + 48\,v_4 - 20\,w_4 \le 0,\quad 45u_4 + 30v_4 - 30w_4 \le 0,$

$120u_4 + 20v_4 - 40w_4 \le 0,\quad$ and u_4, v_4, and $w_4 \ge 0$.

▪ **Excel-Solver** for the efficiency of Branch #4:

	A	B	C	D	E	F	G	H
1	Branch	Personal	Business	Staff	Weighted output	Weighted input	Efficiency	Constraint
2	#1	30	90	30	1.000	1.000	1.000	0.000
3	#2	120	45	30	1.000	1.000	1.000	0.000
4	#3	16	48	20	0.533	0.667	0.800	-0.133
5	#4	45	30	30	**0.500**	1.000	0.500	-0.500
6	#5	120	20	40	0.762	1.333	0.571	-0.571
7	Weights=	0.005	0.010	0.033				

- Objective function: Cell **E5**
- Decision variables Cells **B7:D7**
- Constraints **F5** = 1, **H2:H6** $\leq$ 0, **B7:D7** > 0.01

▪ **LP solution**

The optimal weights are u_4=0.005, v_4=0.010, w_4=0.033. z_4=0.5; That is, the efficiency of Branch #4 is **50%**.

▪ **Final results**

Branch i	u_i	v_i	w_i	Weighted output	Weighted input	Efficiency
#1	0.00478	0.00946	0.0333	1.00	1.00	1.00
#2	0.00478	0.00946	0.0333	1.00	1.00	1.00
#3	0.00714	0.01429	0.0500	0.80	1.00	0.80
#4	0.00478	0.00946	0.0333	0.50	1.00	0.50
#5	0.00625	0.0000	0.0250	0.75	1.00	0.75

- If one more *personal* (or *business*) transaction is processed, the efficiency is increased by u_i (or by v_i) at Branch #*i*.
- Since v_i is twice of u_i, a *business* transaction is twice more important than a *personal* transaction. (Note that the slope of the efficient frontier is -1/2.)
- Value judgment: If one business transaction is worth at least 4 personal transactions, then add a constraint, $v_4 \geq 4u_4$.

* Strengths of DEA

- *DEA* can handle multiple input and multiple output models.
- It doesn't require an assumption of a functional form relating inputs to outputs.
- *DMUs* are directly compared against a peer or combination of peers.
- Inputs and outputs can have very different units.

* Limitations of DEA

- Since *DEA* is an extreme point technique, noise (even symmetrical noise with zero mean) such as measurement error can cause significant problems.
- *DEA* estimates "relative" efficiency of a *DMU*, but it converges very slowly to "absolute" efficiency. In other words, it can tell you how well you are doing compared to your peers, but not compared to a "theoretical maximum."
- Since *DEA* is a *nonparametric* technique, statistical hypothesis tests are difficult and are the focus of ongoing research.
- Since a standard formulation of *DEA* creates a separate linear program for each *DMU*, large problems can be computationally intensive.
- Constant returns to scale (*CRS*) means that the producers are able to *linearly* scale the inputs and outputs without increasing or decreasing efficiency. This is a significant assumption. The assumption of *CRS* may be valid over limited ranges but its use must be justified.

C. Goal Programming*

* Assumptions

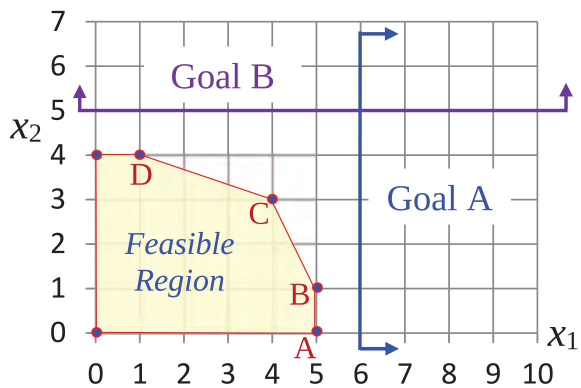
1. The set of attributes 1, 2, …, n is *mutually preferentially independent* so that the decision maker's preferences could be represented by an *additive* value (or cost) function.
2. No uncertainty is present so that all the parameter values are known.

* Procedure

- Formulate as an LP with *deviation* (or *deviational*) *variables* for each constraint and the *weight* for each goal.

Ex] Goal programming

- Constraints:
  Feasible region
- Goal A:
  $x_1 \geq 6$
- Goal B:
  $x_2 \geq 5$



- Deviational variables:

Goal A $x_1 + a_1 - b_1 = 6$
Goal B $x_2 + a_2 - b_2 = 5$

Extreme points	(x_1, x_2)	Deviations a_1	a_2	If the objective function is $z=a_1+a_2$	$z=3a_1+a_2$	$z=a_1+4a_2$
A	(5, 0)	1	5	6	8	21
B	(5, 1)	1	4	5	7	17
C	(4, 3)					
D	(1, 4)	5	1	6	16	9

Ex 1] TV Advertising Problem

Madonna Advertising Agency is trying to determine a TV advertising schedule for Ford Auto Company. Ford has 3 goals: Its ads should be seen by

Goal	Penalty/million
1: At least 60 million high-income men (HIM)	200
2: At least 24 million low-income people (LIP)	100
3: At least 154 million high-income women (HIW)	50

Madonna can purchase two types of ads: (1) Ads shown during soap operas and (2) ads shown during football games. At most $800,000 can be spent on ads. The advertising costs and potential audiences of a one-minute ad of each type are shown in Table. Madonna must determine how many soap opera ads and football ads to purchase for Ford.

	HIM	LIP	HIW	Cost/minute
Soap opera ad	4 million	12 million	14 million	$80,000
Football ad	15 million	2 million	11 million	$100,000

- Variables

  x_1 = number of minutes of ads shown during soap operas

  x_2 = number of minutes of ads shown during football games

- Functional constraints

- Goals

  - Goal 1
  - Goal 2
  - Goal 3

▪ Deviation variables

a_j = amount by which we numerically under the jth goal

b_j = amount by which we numerically exceed the jth goal

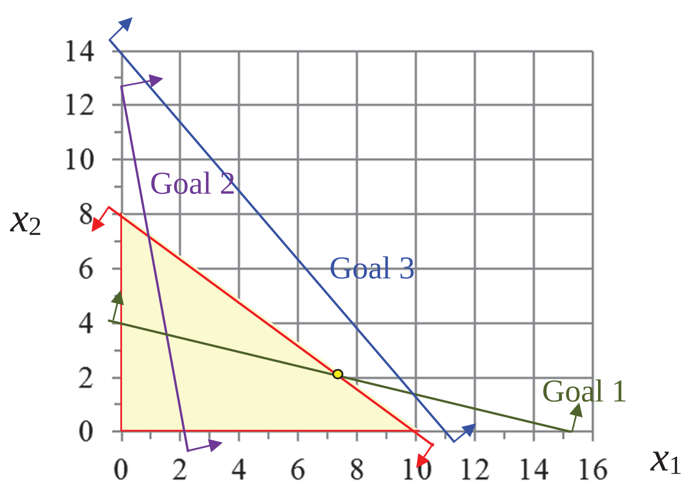
▪ Goal constraints with *deviation* variables

- Goal 1
- Goal 2
- Goal 3
- Constraint
- All variables non-negative.

▪ Weights in the objective function

The total penalty from lost sales caused by deviation from the three goals:

Min z =



▪ Optimal solution is (x_1=7.5, x_2 = 2) with

(a_1 = 0, b_1 = 0), (a_2 = 0, b_2 = 70), and (a_3 = 27, b_3 = 0)

Ex 2] Linear Regression Model

On a slow day, a realtor wondered whether the prices of the houses in her town that had recently sold might follow a simple linear model. She guessed that the sales price of a home (y) might depend linearly on the size (x_1) and age (x_2) of the home as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \qquad \text{where } i = 1, 2, \ldots, n.$$

Find the regression parameters that minimize the sum of the absolute values of the errors ε_i.

- Variables

b_j: Regression parameters, where j = 0, 1 and 2. (They are *unrestricted* in sign.)

(a) Model I

d_i^+ and d_i^-: *Positive* and *negative* deviations.

$Min \quad z = \sum_{i=1}^{n}(d_i^+ + d_i^-)$ subject to

$y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} = d_i^+ - d_i^-$, and $d_i^+, d_i^- \geq 0$.

If the penalty function is $Min\ z = \sum_{i=1}^{n}(2d_i^+ + 5d_i^-)$?

(b) Model II

d_i: Absolute errors, where i = 1, 2, …, n.

$Min \quad z = \sum_{i=1}^{n} d_i$ subject to

$-d_i \leq y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} \leq d_i$, and $d_i \geq 0$.

How to minimize the *maximum* absolute value?