Southeast University
School of Transportation

Urban Transportation Planning

Chinese-English course (2019)

Lecturer: Dr. Qiong Bao

Email: baoqiong@seu.edu.cn

Phone: 17351788445

9:50AM, Friday, 24th May

School of Transportation | Southeast University of China | Campus Jiulonghu | Building Jizhong Y311

Review of uncertainty analysis

- Uncertainty analysis:
  - Input uncertainty
    - Coefficient of variation
    - Effects of population fraction size on uncertainty
  - Model uncertainty
    - Stochastic error
      - Confidence Interval
    - Impact of specific zonal characteristics

Lecture schedule

Lecture	Week	Date/Time	Topic
1	9	28 April 9：50-12：15	Transportation planning & demand and supply & trip-based model
2	10	5 May 9：50-12：15	ABM：data process
3	11	10 May 9：50-12：15	ABM：scheduling
4	12	17 May 9：50-12：15	ABM：uncertainty analysis
5	13	24 May 9：50-12：15	ABM：sensitivity analysis
6	14	31 May 9：50-12：15	Project Evaluation I
7	15	7 June 9：50-12：15	Festival
8	16	14 June 9：50-12：15	Project Evaluation II

Outline

- Uncertainty analysis:

  – to quantify the uncertainty around the mean estimate of one or more outcomes.

- Sensitivity analysis:

  – to identify the most significant parameters in the model and to quantify how the input uncertainty influences the outputs.

    - Local approach
      - one-at-a-time sensitivity analysis measure
    - Global approach
      - the improved Sobol’ method

UA vs. SA

- In general, uncertainty analysis estimates the uncertainty in the output taking into account the uncertainty affecting the input factors.
- Rather than being a unique value the estimated output represents a distribution of values and elementary statistics such as the mean, standard deviation and percentiles are used to describe its features.
- The purpose of uncertainty analysis is to quantify the uncertainty around the mean estimate of one or more outcomes.

UA vs. SA

- Sensitivity analysis is defined as the study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input.
- The objective of the sensitivity analysis is to identify the most significant parameters in the model and to quantify how the input uncertainty influences the outputs.
- In other words, the sensitivity analysis could also identify the impact of input variables on the choice alternatives.

Sensitivity analysis principle

- In particular, by performing the sensitivity analysis, the considered model is evaluated a specified number of times with different values of the input variables. Based on the results, a better understanding of the influence of the different input variables and their variations on the model outputs can be provided.
- In short, a thorough sensitivity analysis helps to interpret the model, increases its credibility across a range of input scenarios and can uncover underlying errors.

Sensitivity analysis approach

- Sensitivity analysis techniques consist mainly of local approaches and global approaches.

- A local approach: addresses sensitivity relative to point estimates of parameter values, in which inputs are varied one at a time by a small amount around some fixed point and the effects of individual variation on the output are calculated.

- A global approach: evaluates the effect of a parameter while all other parameters are varied as well and thus the entire effect on the output and interactions between input parameters can be assessed.

Local sensitivity analysis approach

- A local sensitivity analysis approach:
  - inputs varied one at a time and keeping all other variables as observed;
  - easy to perform;
  - need no detailed knowledge of variable distribution;
  - inefficient when the number of variables is large;
  - cannot take into account interactions between multiple variables.
- Typical local sensitivity analysis approach:
  - such as partial derivative method, one-at-a-time sensitivity measures method, and the sensitivity index method

Global sensitivity analysis approach

- A global sensitivity analysis approach:
  - evaluate the effect of an input while all other variables are varied;
  - assess entire effect on the output and interactions between different input variables.
- Typical global sensitivity analysis approach:
  - for instance, regression based approaches, regionalized sensitivity analysis, the Morris method, Fourier amplitude sensitivity test (FAST) and its extended version (eFAST), as well as the (improved) Sobol' method

Local approach VS. Global approach

In general, each approach has its own advantages and disadvantages.

- In a local SA approach: by varying the input variables one after another, and keeping all other variables as observed. There is no need to estimate the distribution of each input variable. However, the local approach cannot take into account interactions resulting from the simultaneous variation of multiple input variables.

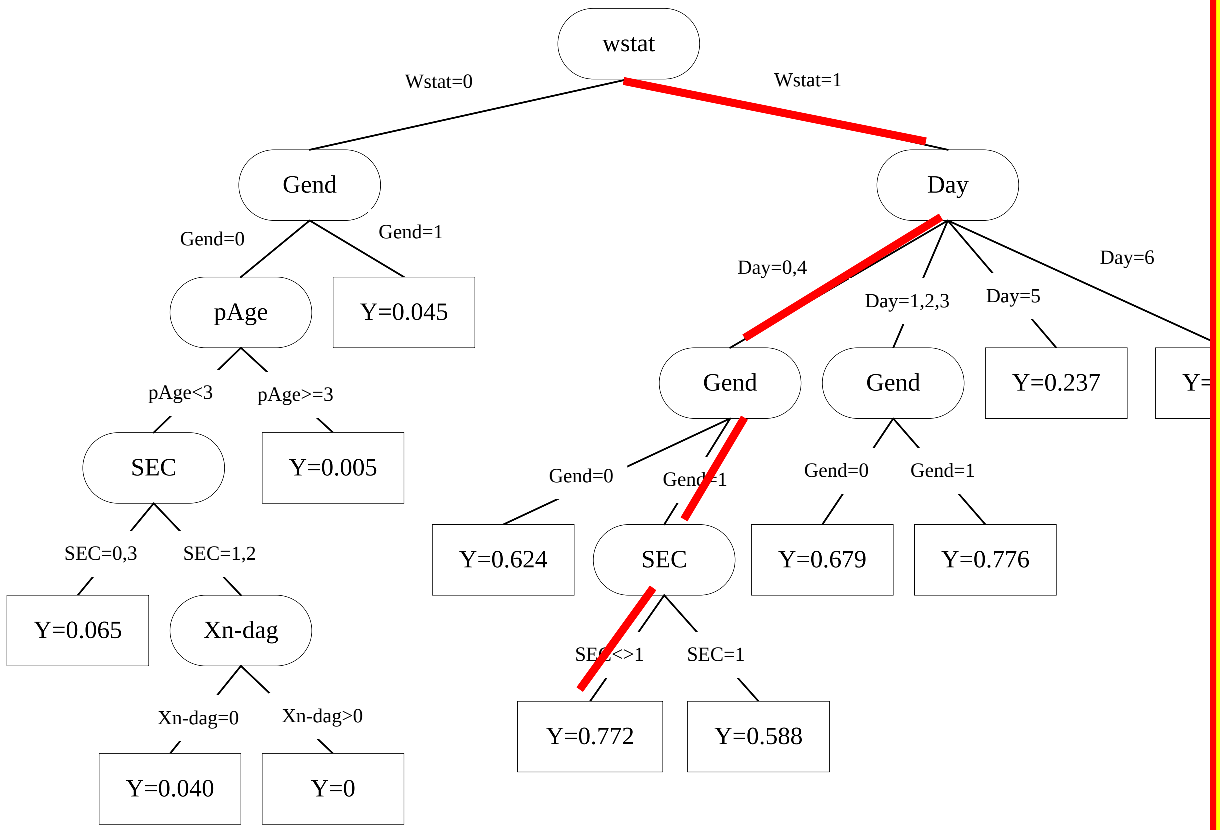
- In a global SA approach: the entire effect on the output and interactions between input variables could be assessed. To do the calculation, a number of computer programs need to be applied. And the input distribution need to be investigated first.

Typical sensitivity analysis approach

In this class, we will discuss both the local and global sensitivity analysis approaches.

- Local approach:
  - the one-at-a-time sensitivity measure
- Global approach:
  - the improved Sobol’ method

DT: work activity choice



DT concerning work activity implementation

Y: the probability of making the final decision, i.e. impl

6 input variables:

- *Wstat=1-- Have a work*
- *Day=0 -- Monday*
- *Gend=1 -- Female*
- *pAge=2 -- [55, 65)*
- *SEC=2 -- [2250, 3250)*
- *Xn-dag=3 -- (762,938]*

Input=[1,0,1,2,…]

Y=0.772

Local approach: one-at-a-time sensitivity measure

- The procedure of OAT sensitivity measure:

  – For the work activity choice decision tree, 6 input condition variables are involved which collectively determine whether a work-related activity will be implemented or not.

  – In order to measure the relative importance of each input variable on the choice variable, we further compute the choice frequency distribution for each input variable by varying the value of the selected input variable and keeping all the others as observed for all possible cases.

  – calculate the Chi-square (χ^2) of the frequency table, together with the sensitivity measures IS_{work} and MS_{work}

Local approach: one-at-a-time sensitivity measure

For the DT concerning work activity implementation:

Condition variable	Definition	Nr. of categories	Condition levels
Wstat	Work status	2	0: no work, 1: work
Day	Day of the week	7	0: Monday … 6: Sunday
Gend	Gender of individual	2	0: male, 1: female
pAge	Age of the person	5	0: <35, 1: [35, 55), 2: [55, 65), 3: [65, 75), 4: >=75
SEC	Income	4	0: [0, 1250), 1: [1250, 2250), 2: [2250, 3250), 3: >=3250
Xn-dag	Number of employees	6	0: (0,395], 1: (395,635], 2: (635,762], 3: (762,938], 4: (938, 2525], 5: >2525

Local approach: one-at-a-time sensitivity measure

- Step1: List the condition variables and the discrete values of each variable.

Condition variable	Nr. of categories	Condition levels
Wstat	2	0,1
Day	7	0,1,2,3,4,5,6
Gend	2	0,1
pAge	5	0,1,2,3,4
SEC	4	0,1,2,3
Xn-dag	6	0,1,2,3,4,5

Local approach: one-at-a-time sensitivity measure

Condition variable	Nr. of categories	Condition levels
Wstat	2	0,1
Day	7	0,1,2,3,4,5,6
Gend	2	0,1

Permutation and Combination:

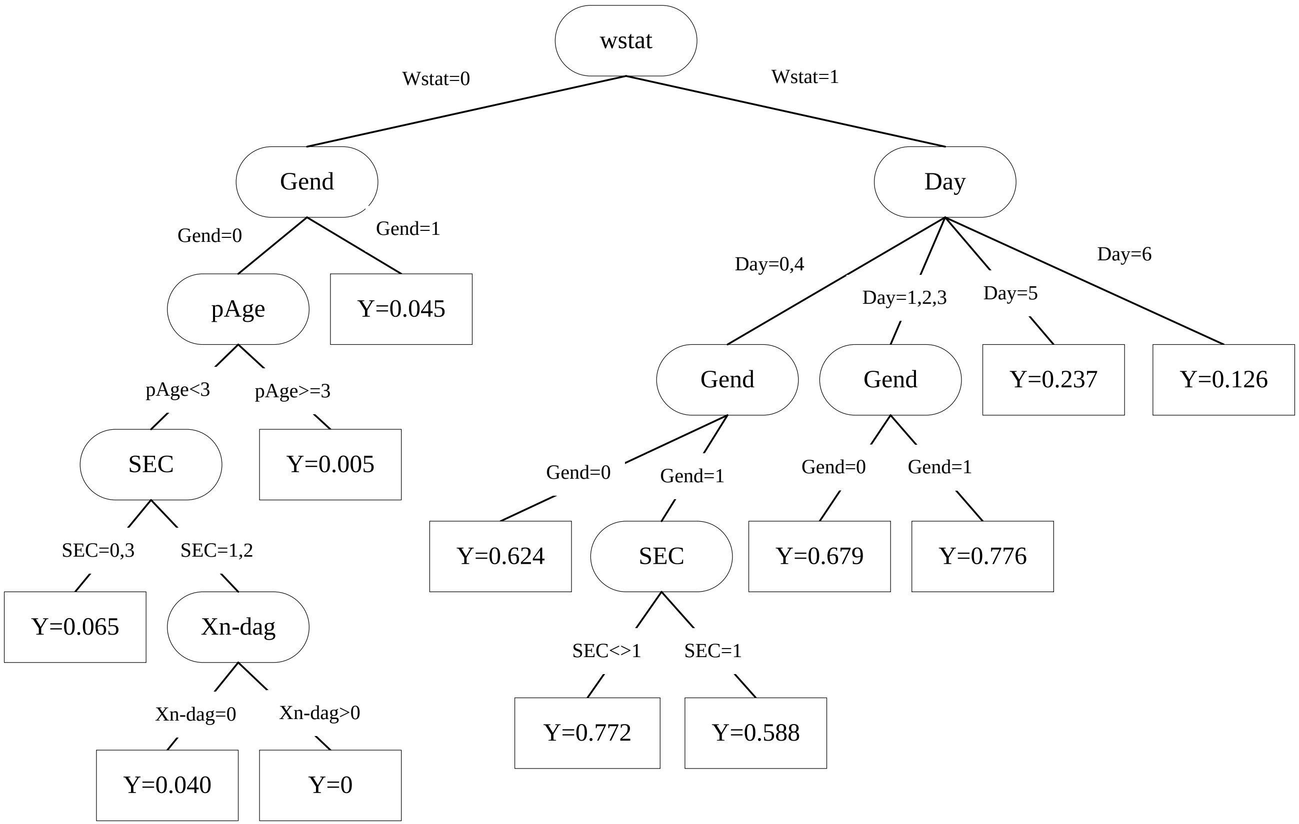
$$P_n^m = \frac{n!}{(n-m)!}$$

$$C_n^m = \frac{n!}{m!\,(n-m)!}$$

- Step2: Consider the nr. of categories for each condition variable and answer *how many combinations of input variables could we have?*
- And calculate the total possible combinations of input variables:

$$\text{Combinations of input} = C_2^1 C_7^1 C_2^1 C_5^1 C_4^1 C_6^1 = 3{,}360$$

Local approach: one-at-a-time sensitivity measure



Condition variable	Condition levels
Wstat	0,1
Day	0,1,2,3,4,5,6
Gend	0,1
pAge	0,1,2,3,4
SEC	0,1,2,3
Xn-dag	0,1,2,3,4,5

- Step3: Based on our constructed DT, calculate the probability of implement work activity for each input possibility (T=3360).

  E.g., input=(0,0,0,0,0,0)=0.065; input=(1,1,0,0,0,0)=0.679

Calculate the probability of working choice (Y) w.r.t. ‘Wstat’

1	Wstat=0						
2	0	0	0	0	0	0	0. 065
3	0	0	0	0	0	1	0. 065
4	0	0	0	0	0	2	0. 065
5	0	0	0	0	0	3	0. 065
6	0	0	0	0	0	4	0. 065
7	0	0	0	0	0	5	0. 065
8	0	0	0	0	1	0	0. 065
9	0	0	0	0	1	1	0. 065
10	0	0	0	0	1	2	0. 065
1667	0	1	4	3	4	3	0. 045
1668	0	1	4	3	4	4	0. 045
1669	0	1	4	3	4	5	0. 045
1670	0	1	4	3	5	0	0. 045
1671	0	1	4	3	5	1	0. 045
1672	0	1	4	3	5	2	0. 045
1673	0	1	4	3	5	3	0. 045
1674	0	1	4	3	5	4	0. 045
1675	0	1	4	3	5	5	0. 045
1676	0	1	4	3	6	0	0. 045
1677	0	1	4	3	6	1	0. 045
1678	0	1	4	3	6	2	0. 045
1679	0	1	4	3	6	3	0. 045
1680	0	1	4	3	6	4	0. 045
1681	0	1	4	3	6	5	0. 045

1	Wstat=2						
1682	2	0	0	0	0	0	0. 624
1683	2	0	0	0	0	1	0. 624
1684	2	0	0	0	0	2	0. 624
1685	2	0	0	0	0	3	0. 624
1686	2	0	0	0	0	4	0. 624
1687	2	0	0	0	0	5	0. 624
1688	2	0	0	0	1	0	0. 679
1689	2	0	0	0	1	1	0. 679
1690	2	0	0	0	1	2	0. 679
1691	2	0	0	0	1	3	0. 679
3348	2	1	4	3	4	4	0. 772
3349	2	1	4	3	4	5	0. 772
3350	2	1	4	3	5	0	0. 237
3351	2	1	4	3	5	1	0. 237
3352	2	1	4	3	5	2	0. 237
3353	2	1	4	3	5	3	0. 237
3354	2	1	4	3	5	4	0. 237
3355	2	1	4	3	5	5	0. 237
3356	2	1	4	3	6	0	0. 126
3357	2	1	4	3	6	1	0. 126
3358	2	1	4	3	6	2	0. 126
3359	2	1	4	3	6	3	0. 126
3360	2	1	4	3	6	4	0. 126
3361	2	1	4	3	6	5	0. 126

Calculate the probability of working choice (Y) w.r.t. 'SEC'

1	SEC=0						
2	0	0	0	0	0	0	0.065
3	0	0	0	0	0	1	0.065
4	0	0	0	0	0	2	0.065
5	0	0	0	0	0	3	0.065
6	0	0	0	0	0	4	0.065
7	0	0	0	0	0	5	0.065
836	2	1	4	0	6	0	0.126
837	2	1	4	0	6	1	0.126
838	2	1	4	0	6	2	0.126
839	2	1	4	0	6	3	0.126
840	2	1	4	0	6	4	0.126
841	2	1	4	0	6	5	0.126

1	SEC=1						
842	0	0	0	1	0	0	0.04
843	0	0	0	1	0	1	0
844	0	0	0	1	0	2	0
845	0	0	0	1	0	3	0
846	0	0	0	1	0	4	0
847	0	0	0	1	0	5	0
1676	2	1	4	1	6	0	0.126
1677	2	1	4	1	6	1	0.126
1678	2	1	4	1	6	2	0.126
1679	2	1	4	1	6	3	0.126
1680	2	1	4	1	6	4	0.126
1681	2	1	4	1	6	5	0.126

1	SEC=2						
1682	0	0	0	2	0	0	0.04
1683	0	0	0	2	0	1	0
1684	0	0	0	2	0	2	0
1685	0	0	0	2	0	3	0
1686	0	0	0	2	0	4	0
1687	0	0	0	2	0	5	0
2516	2	1	4	2	6	0	0.126
2517	2	1	4	2	6	1	0.126
2518	2	1	4	2	6	2	0.126
2519	2	1	4	2	6	3	0.126
2520	2	1	4	2	6	4	0.126
2521	2	1	4	2	6	5	0.126

1	SEC=3						
2522	0	0	0	3	0	0	0.065
2523	0	0	0	3	0	1	0.065
2524	0	0	0	3	0	2	0.065
2525	0	0	0	3	0	3	0.065
2526	0	0	0	3	0	4	0.065
2527	0	0	0	3	0	5	0.065
3356	2	1	4	3	6	0	0.126
3357	2	1	4	3	6	1	0.126
3358	2	1	4	3	6	2	0.126
3359	2	1	4	3	6	3	0.126
3360	2	1	4	3	6	4	0.126
3361	2	1	4	3	6	5	0.126

Local approach: one-at-a-time sensitivity measure

Condition variable	Definition	Nr. of categories	Condition levels
Wstat	Work status	2	0: no work, 1: work
Day	Day of the week	7	0: Monday … 6: Sunday
Gend	Gender of individual	2	0: male, 1: female
pAge	Age of the person	5	0: <35, 1: [35, 55), 2: [55, 65), 3: [65, 75), 4: >=75
SEC	Income	4	0: [0, 1250), 1: [1250, 2250), 2: [2250, 3250), 3:
Xn-dag	Number of employees		5,762], 3: (762,938], 4:

Wstat=0, overall probability=57.54

Wstat=1, overall probability=934.92

- Step4: Calculate the overall probability of implement work activity for each condition level w.r.t. each input variable.

Calculate the overall probability w.r.t. 'Wstat'

1	Wstat=0						
2	0	0	0	0	0	0	0. 065
3	0	0	0	0	0	1	0. 065
4	0	0	0	0	0	2	0. 065
5	0	0	0	0	0	3	0. 065
6	0	0	0	0	0	4	0. 065
7	0	0	0	0	0	5	0. 065
8	0	0	0	0	1	0	0. 065
9	0	0	0	0	1	1	0. 065
10	0	0	0	0	1	2	0. 065
1667	0	1	4	3	4	3	0. 045
							45
							45
							45
							45
1672	0	1	4	3	5	2	0. 045
1673	0	1	4	3	5	3	0. 045
1674	0	1	4	3	5	4	0. 045
1675	0	1	4	3	5	5	0. 045
1676	0	1	4	3	6	0	0. 045
1677	0	1	4	3	6	1	0. 045
1678	0	1	4	3	6	2	0. 045
1679	0	1	4	3	6	3	0. 045
1680	0	1	4	3	6	4	0. 045
1681	0	1	4	3	6	5	0. 045

Wstat=0, overall probability=57.54

1	Wstat=2						
1682	2	0	0	0	0	0	0. 624
1683	2	0	0	0	0	1	0. 624
1684	2	0	0	0	0	2	0. 624
1685	2	0	0	0	0	3	0. 624
1686	2	0	0	0	0	4	0. 624
1687	2	0	0	0	0	5	0. 624
1688	2	0	0	0	1	0	0. 679
1689	2	0	0	0	1	1	0. 679
1690	2	0	0	0	1	2	0. 679
1691	2	0	0	0	1	3	0. 679
							2
							2
							7
							7
3352	2	1	4	3	5	2	0. 237
3353	2	1	4	3	5	3	0. 237
3354	2	1	4	3	5	4	0. 237
3355	2	1	4	3	5	5	0. 237
3356	2	1	4	3	6	0	0. 126
3357	2	1	4	3	6	1	0. 126
3358	2	1	4	3	6	2	0. 126
3359	2	1	4	3	6	3	0. 126
3360	2	1	4	3	6	4	0. 126
3361	2	1	4	3	6	5	0. 126

Wstat=2, overall probability=934.92

Calculate the overall probability w.r.t. ‘SEC’

1				SEC=0			
2	0	0	0	0	0	0	0. 065
3	0	0	0	0	0	1	0. 065
4	0	0	0	0	0	2	0. 065
5	0	0	0	0	0	3	0. 065
6	0	0	0	0	0	4	0. 065
							065
							26
							26
							26
839	2	1	4	0	6	3	0. 126
840	2	1	4	0	6	4	0. 126
841	2	1	4	0	6	5	0. 126

SEC=0, overall probability=254.55

1				SEC=1			
842	0	0	0	1	0	0	0. 04
843	0	0	0	1	0	1	0
844	0	0	0	1	0	2	0
845	0	0	0	1	0	3	0
846	0	0	0	1	0	4	0
							0
							26
							26
							26
1679	2	1	4	1	6	3	0. 126
1680	2	1	4	1	6	4	0. 126
1681	2	1	4	1	6	5	0. 126

SEC=1, overall probability=236.16

1				SEC=2			
1682	0	0	0	2	0	0	0. 04
1683	0	0	0	2	0	1	0
1684	0	0	0	2	0	2	0
1685	0	0	0	2	0	3	0
							0
							0
							26
							26
							26
2519	2	1	4	2	6	3	0. 126
2520	2	1	4	2	6	4	0. 126
2521	2	1	4	2	6	5	0. 126

SEC=2, overall probability=247.2

1				SEC=3			
2522	0	0	0	3	0	0	0. 065
2523	0	0	0	3	0	1	0. 065
2524	0	0	0	3	0	2	0. 065
2525	0	0	0	3	0	3	0. 065
							65
							65
							26
							26
							0. 126
3359	2	1	4	3	6	3	0. 126
3360	2	1	4	3	6	4	0. 126
3361	2	1	4	3	6	5	0. 126

SEC=3, overall probability=254.55

Local approach: one-at-a-time sensitivity measure

- The overall probability of implement work activity for each condition level w.r.t. each input variable.

Overall probability of conducting a work activity							
Condition level	0	1	2	3	4	5	6
wstat	57.54	934.92	——	——	——	——	——
Gend	457.5	534.96	——	——	——	——	——
pAge	200.564	200.564	200.564	195.384	195.38	——	——
SEC	254.55	236.16	247.2	254.55	——	——	——
Day	170.22	182.82	182.82	182.82	170.22	65.1	38.46
Xn_dag	166.81	165.13	165.13	165.13	165.13	165.13	——

Local approach: one-at-a-time sensitivity measure

- Step5: Derive a frequency table for each input variable.

	Work activity	Non-work activity	Row total
Wstat=0	57.54	1622.46	1680
Wstat=1	934.92	745.08	1680
Column total	992.46	2367.54	3360

	Work activity	Non-work activity	Row total
SEC=0	254.55	585.45	840
SEC=1	236.16	603.84	840
SEC=2	247.2	592.8	840
SEC=3	254.55	585.45	840
Column total	992.46	2367.54	3360

- Step6: Calculate the Chi-square (χ^2) of the frequency table.

Example of Chi-square (χ^2)

- Chi-square is used to test for association between categorical variables.

Example: Is a drug effective at curing a disease or not?

The observed patients who received
the drug/placebo were cured (not cured)

	Cured	Not cured
Drug	30	13
Placebo	11	30

Example of Chi-square (χ^2)

- 1. Calculate the row totals, column totals, and grand total for the observed data.

	Cured	Not cured	Row Total
Drug	30	13	43
Placebo	11	30	41
Column Total	41	43	84

- 2. Calcul

(row total *

$$Expected = \frac{row\ total \times column\ total}{grand\ total} = \frac{43 \times 41}{84} = 20.99$$

	Cured	Not cured	Row Total
Drug	20.99	22.012	43
Placebo	20.012	20.99	41
Column Total	41	43	84

Example of Chi-square (χ^2)

- 3. Calculate chi-square value

Observed	Cured	Not cured	Row Total
Drug	30	13	43
Placebo	11	30	41
Column Total	41	43	84

$$\chi^2 = \sum_{i=1}^{n} (c_i - Ec_i)^2 \Big/ Ec_i$$

$$= \frac{(30 - 20.99)^2}{20.99} + \frac{(13 - 22.012)^2}{22.012}$$

$$+ \frac{(11 - 20.012)^2}{20.012} + \frac{(30 - 20.99)^2}{20.99}$$

$$= 15.48$$

Expected	Cured	Not cured	Row Total
Drug	20.99	22.012	43
Placebo	20.012	20.99	41
Column Total	41	43	84

Local approach: one-at-a-time sensitivity measure

- Step6: Calculate the Chi-square ((χ^2) denoted as IS in OAT SA) of the frequency table of work activity choice DT.

Frequency table	Work activity	Non-work activity	Row total
Wstat=0	57.54	1622.46	1680
Wstat=1	934.92	745.08	1680
Column total	992.46	2367.54	3360
Expect value	**496.23**	**1183.77**	
χ^2	**1100.79**		

$$Expected1 = 496.23 = 992.46 \times 1680/3360$$
$$Expected2 = 1183.77 = 2367.54 \times 1680/3360$$

$$\chi^2 = 1100.79 = \frac{(57.54 - 496.23)^2}{496.23} + \frac{(1622.46 - 1183.77)^2}{1183.77} + \frac{(934.92 - 496.23)^2}{496.23} + \frac{(745.08 - 1183.77)^2}{1183.77}$$

Local approach: one-at-a-time sensitivity measure

	Work activity	Non-work activity	Row total
SEC=0	254.55	585.45	840
SEC=1	236.16	603.84	840
SEC=2	247.2	592.8	840
SEC=3	254.55	585.45	840
Column total	992.46	2367.54	3360
Expect value	248.115	591.885	
χ^2	1.296		

$$Expected1 = 248.115 = 992.46 \times 840/3360$$
$$Expected2 = 591.885 = 2367.54 \times 840/3360$$

$$\chi^2 = 1.296 = \sum_{i=1}^{n} (c_i - Ec_i)^2 \Big/ Ec_i$$

Local approach: one-at-a-time sensitivity measure

DT: work activity choice

Condition variable	IS (χ^2)
wstat	1100.79
Day	232.48
Gend	8.58
SEC	1.296
pAge	0.23
Xn_dag	0.02

- The sensitivity measure *IS:*

  *(χ^2 is denoted as **IS** in OAT SA)*

  – *Indicate the overall impact of the condition variable on the choice variable.*

  – *A higher value of IS implies a more important condition variable for the choice alternative.*

  - *'wstat' is the most important variable for work activity choice, followed by 'Day'.*
  - *Although the remaining variables also have some impact on the choice, their importance level is much lower.*

Global approach: the improved Sobol' method

- A global sensitivity analysis approach, evaluates the effect of a parameter while all other parameters are varied as well.
- And thus the entire effect on the output and interactions between input parameters can be assessed.
- Moreover, the global sensitivity method can be applied to arbitrary nonlinear functions.
- The Sobol' method, originally proposed by Sobol' (1990), is a variance-based method.
- The Sobol' method allows for simultaneous variation of the values of all input variables, in contrast to the simple one-at-a-time sensitivity analysis.

Global approach: the improved Sobol' method

- Assume a model output ***Y*** can be written as a function of its input variables $\boldsymbol{x_1}$ to $\boldsymbol{x_k}$.

$$Y = f(X) = f(x_1, x_2, \cdots, x_k)$$

- Each input variable has a range of variation that might lead to some uncertainty of the model output. So **the total variance of *Y*** can be expressed as follows,

$$V(Y) = \sum_{i=1}^{k} V_i + \sum_{i=1}^{k} \sum_{j>i}^{k} V_{ij} + \cdots + V_{12 \cdots k}$$

Global approach: the improved Sobol' method

$$\boldsymbol{V(Y)} = \sum_{i=1}^{k} \boldsymbol{V_i} + \sum_{i=1}^{k}\sum_{j>i}^{k} \boldsymbol{V_{ij}} + \cdots + \boldsymbol{V_{12\cdots k}}$$

$$V_i = V\left[E\left(Y \mid X_i = x_i^*\right)\right]$$

$$V_{ij} = V\left[E\left(Y \mid X_i = x_i^*, X_j = x_j^*\right)\right] - V_i - V_j$$

Where,

$\boldsymbol{V_i}$ is the main effect of $\boldsymbol{X_i}$ on $\boldsymbol{Y}$ given that $\boldsymbol{X_i}$ has a fixed value X_i^*.

$\boldsymbol{V_{ij}}$ is the joint effect of the pair ($\boldsymbol{X_i}$, $\boldsymbol{X_j}$) on $\boldsymbol{Y}$ given that the inputs $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$ have fixed values X_i^* and X_j^*, respectively.

Global approach: the improved Sobol' method

- Then we can calculate the first-order sensitivity index $\boldsymbol{S_i}$

$$\boldsymbol{S_i = \frac{V_i}{V(Y)} = \frac{V[E(Y|X_i)]}{V(Y)}}$$

- and the total-effect sensitivity index $\boldsymbol{S_{T_i}}$ for a given input $\boldsymbol{X_i}$.

$$\boldsymbol{S_{T_i} = S_i + \sum_{j \neq i} S_{ij} + \cdots + S_{12\cdots k} = 1 - \frac{V[E(Y|X_{\sim i})]}{V(Y)}}$$

$\boldsymbol{X_{\sim i}}$ denotes all of the input variables other than $\boldsymbol{X_i}$.

- The first-order sensitivity index $\boldsymbol{S_i}$: quantify the individual impact of input variable.

- The total-effect sensitivity index $\boldsymbol{S_{T_i}}$: accounts for the total contribution from the input variable.

Global approach: the improved Sobol' method

- In other words, the first-order sensitivity index S_i quantifies the effect of varying X_i alone.
- And the total-effect sensitivity index S_{T_i} includes the variance derived from X_i and also from its any combination with the other variables.
- It is effective, while the main drawback of applying the standard sobol' method is its computational cost. So in 2002, Saltelli proposed an improved Sobol' method, which is a Monte-Carlo based implementation. In this approach, the first-order and the total effect sensitivity indices are calculated based on three input variable sampling matrices. The improved Sobol' method is developed for a faster calculation.

Global approach: the improved Sobol' method

- The important features of the improved Sobol' method are:
  - First, the model is independence. That is, the sensitivity measure is model-free and thus it could be applied to identify the most influential factors even when the model is complex or unknown;
  - Moreover, it is a global method capable of capturing the influence of each input factor on the full range of output variation. That is, the total effect index accounts for how the variance of a certain output depends not only on variations of the single input (first-order effect), but also on its interaction effects with the other inputs (higher-order effects).
  - In addition, the interpretation of the results is very intuitive and straightforward.

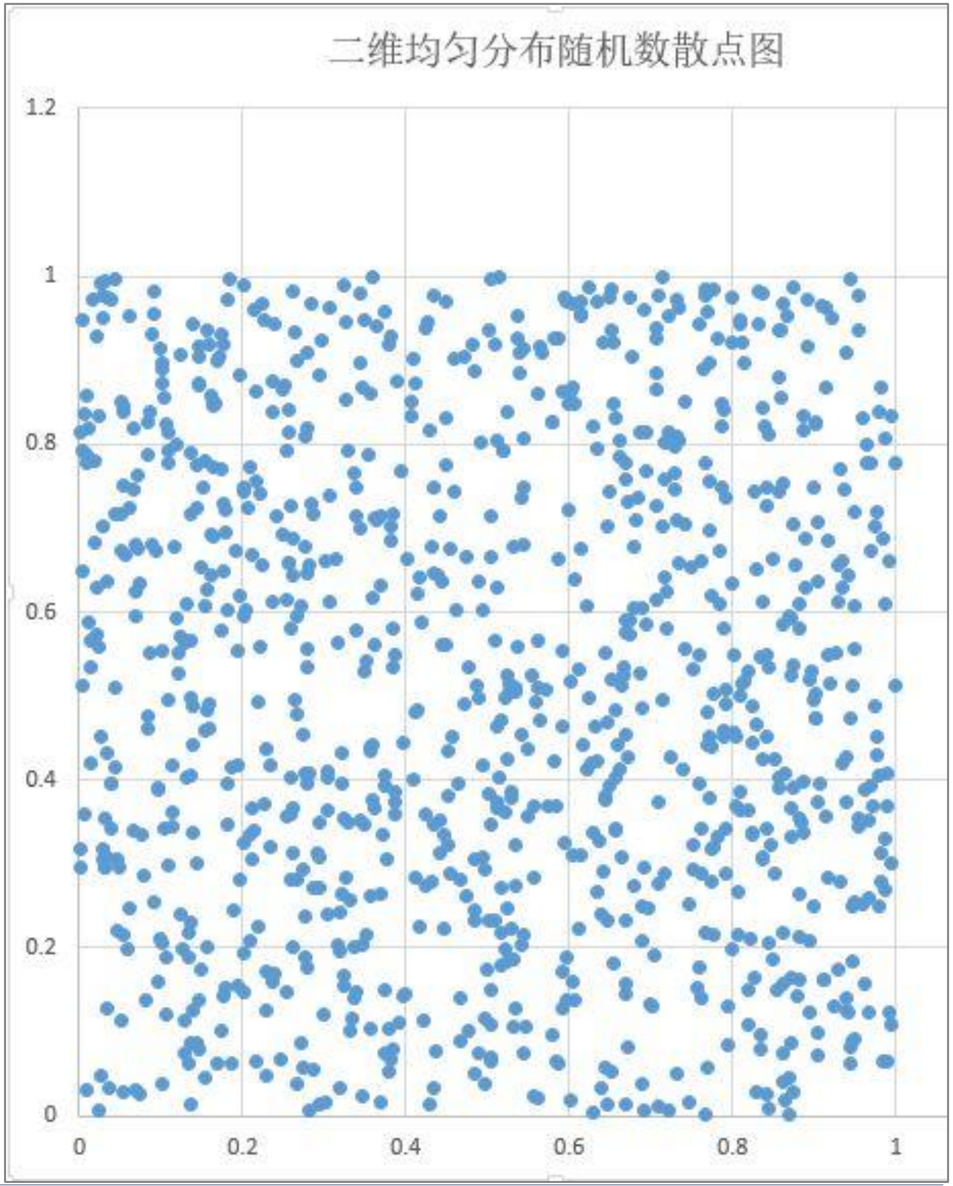
Global approach: the improved Sobol' method

- Steps of the improved Sobol' method:
  - Step1: We first create two independent input variable sampling matrices ***A*** and ***B*** with dimension (N, k), where N is the sample size and k is the number of input variables. Thus, each row in matrices ***A*** and ***B*** represents a possible value of ***X***.

$$A = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_i^{(1)} & \cdots & x_k^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_i^{(2)} & \cdots & x_k^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_1^{(N-1)} & x_2^{(N-1)} & \cdots & x_i^{(N-1)} & \cdots & x_k^{(N-1)} \\ x_1^{(N)} & x_2^{(N)} & \cdots & x_i^{(N)} & \cdots & x_k^{(N)} \end{bmatrix} \quad B = \begin{bmatrix} x_{k+1}^{(1)} & x_{k+2}^{(1)} & \cdots & x_{k+i}^{(1)} & \cdots & x_{2k}^{(1)} \\ x_{k+1}^{(2)} & x_{k+2}^{(2)} & \cdots & x_{k+i}^{(2)} & \cdots & x_{2k}^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{k+1}^{(N-1)} & x_{k+2}^{(N-1)} & \cdots & x_{k+i}^{(N-1)} & \cdots & x_{2k}^{(N-1)} \\ x_{k+1}^{(N)} & x_{k+2}^{(N)} & \cdots & x_{k+i}^{(N)} & \cdots & x_{2k}^{(N)} \end{bmatrix}$$

Random sequence

- Random number
  - Generating random number between 0 and 1
  - When computer continuously produces lots of random, there will appear a phenomenon that the later random sequence is the same with the former one, that is Random is not really random.



Random sequence

- Pseudorandom number
  - The random number produced on the computer by mathematical method
- low-discrepancy sequence
  - E.g., Halton sequence, Faure sequence, Sobol sequence
  - Principle: the generated random numbers are more uniformly filled with unit hypercubes.
  - Subinterval: the probability that the uniform random number generated on [0,1] falls on the subintervals is the same.

Random sequence

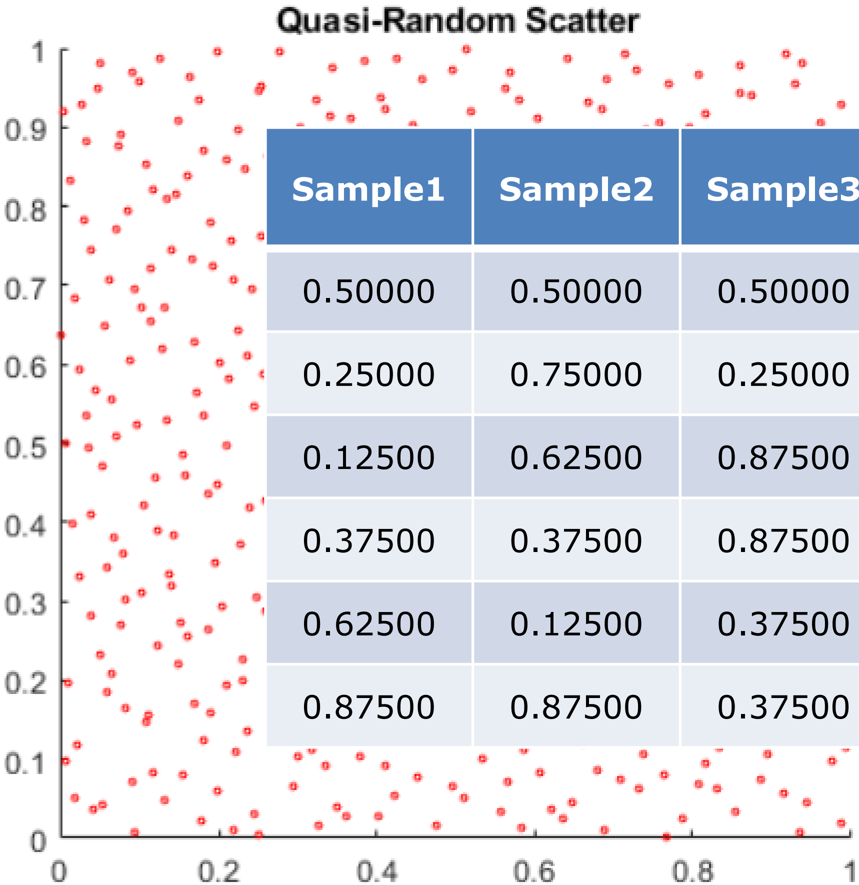
- Steps of the improved Sobol’ method:

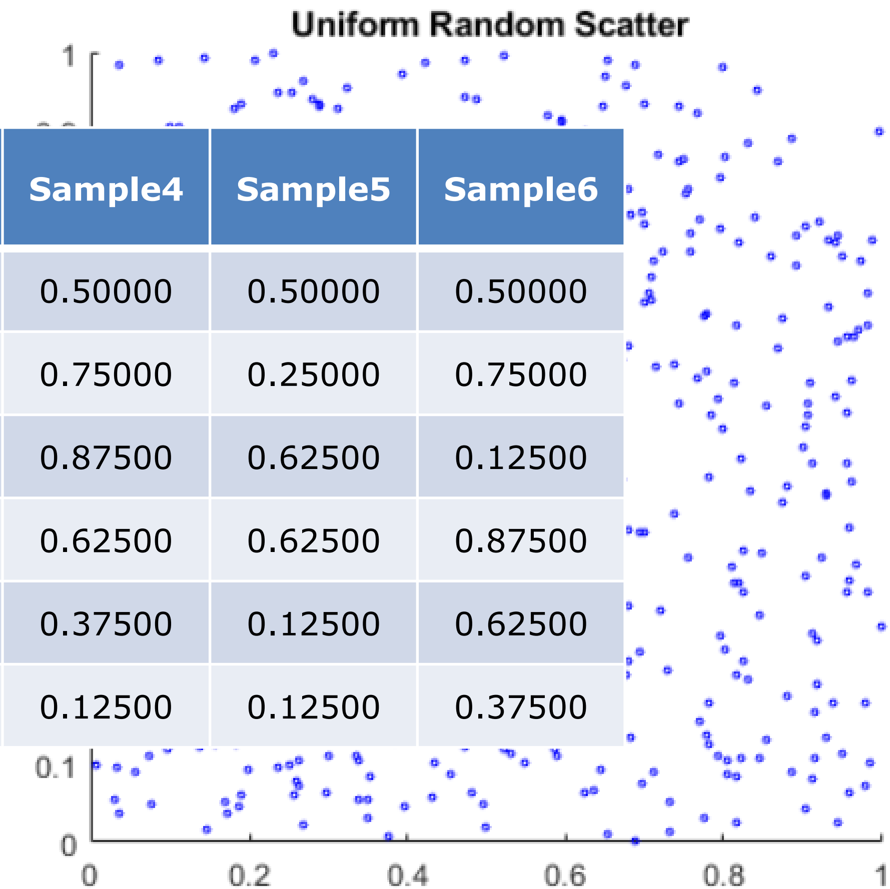
  - In practice, to make the samples more homogeneously distributed in the whole range of variability of input variables, the Sobol’ quasi-random sequence (Sobol’, 1976) is selected with a size of (N, 2k), where N is the sample size and k is the number of input variables. Thus, all the sampling points are uniformly distributed in the space of (0,1).

  - Quasi-random sequence is also known as low-discrepancy sequence. The sequences seek to fill space uniformly and use a base of 2 to form successively finer uniform partitions of the unit interval, and then reorder the coordinates in each dimension.

Random sequence

- The quasi-random scatter appears more uniform, avoiding the clumping in the pseudorandom scatter.





Sample1	Sample2	Sample3	Sample4	Sample5	Sample6
0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
0.25000	0.75000	0.25000	0.75000	0.25000	0.75000
0.12500	0.62500	0.87500	0.87500	0.62500	0.12500
0.37500	0.37500	0.87500	0.62500	0.62500	0.87500
0.62500	0.12500	0.37500	0.37500	0.12500	0.62500
0.87500	0.87500	0.37500	0.12500	0.12500	0.37500

Reference

- low-discrepancy sequence
  - J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluation multi-dimensional integrals. Math. 1960, (2): 84-90.
  - H. Niederreiter. Quasi-monte carlo methods and pseudo-random numbers [J]. Bull. Amer. Math. Phys. Sos. 1978,84 (6):957-1041.
  - LM. Sobol. The distribution of points in a cube and the approximate evaluation of integrals [J]. USSR Comp. Math. And Math. Plays. 1967, (7): 86-112.
  - Galanti S. Jun A. Low-discrepancy Sequences: Monte Carlo Simulation of Option Prices [J]. Journal of Derivatives. Fall 1997: 63-83.

Global approach: the improved Sobol' method

- Steps of the improved Sobol' method:
  - Step2: we define a matrix $\boldsymbol{C_i}$ formed by all columns of $\boldsymbol{B}$ except the $\boldsymbol{i_{th}}$ column, which is taken from A:

$$C_i = \begin{bmatrix} x_{k+1}^{(1)} & x_{k+2}^{(1)} & \cdots & x_i^{(1)} & \cdots & x_{2k}^{(1)} \\ x_{k+1}^{(2)} & x_{k+2}^{(2)} & \cdots & x_i^{(2)} & \cdots & x_{2k}^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{k+1}^{(N-1)} & x_{k+2}^{(N-1)} & \cdots & x_i^{(N-1)} & \cdots & x_{2k}^{(N-1)} \\ x_{k+1}^{(N)} & x_{k+2}^{(N)} & \cdots & x_i^{(N)} & \cdots & x_{2k}^{(N)} \end{bmatrix} \quad i = 1, 2, \ldots k$$

Global approach: the improved Sobol' method

- Steps of the improved Sobol' method:
  - Step3: We now compute the model output for all the input values in the sample matrices ***A***, ***B***, and $\boldsymbol{C_i}$, obtaining three vectors of model outputs with a dimension of $N \times 1$.

$$y_A = f(A) \qquad y_B = f(B) \qquad y_{C_i} = f(C_i)$$

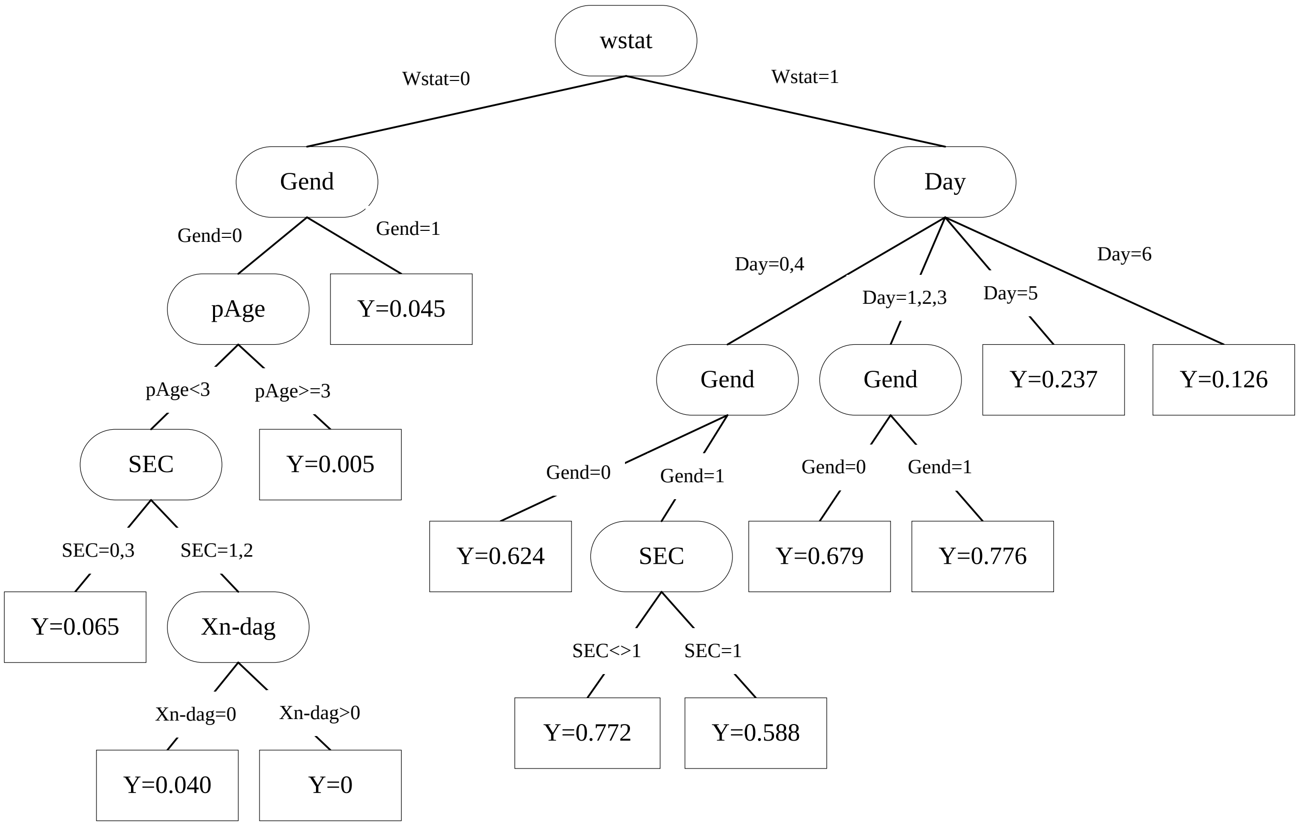
Sample1	Sample2	Sample3	Sample4	Sample5	Sample6	Y
0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.04000
0.25000	0.75000	0.25000	0.75000	0.25000	0.75000	0.04500
0.12500	0.62500	0.87500	0.87500	0.62500	0.12500	0.04500
0.37500	0.37500	0.87500	0.62500	0.62500	0.87500	0.00500
0.62500	0.12500	0.37500	0.37500	0.12500	0.62500	0.62400
0.87500	0.87500	0.37500	0.12500	0.12500	0.37500	0.77200

Global approach: the improved Sobol' method

```
% x1, x2, x3, x4, x5, x6    [0-1] wstat,

% Input data for decision tree

%wstat                  [0 = no work, 2 = work]
  if x1 <= 0.5
    wstat = 0;
else
    wstat = 2;
end

%Gend               [0 = male, 1 = female]
  if x2 <= 0.5
    Gend = 0;
else
    Gend = 1;
end

%pAge              [pAge = 0 - 2; 3 - 4]
if x3 <= 0.8
    pAge = 0;
else
    pAge = 3;
end
```

...on of ...ation		Mean	Std. Dev.	Distribution of DT
	0.5	1.05	1	0.437
	1.0			0.563
	0.5	1.52	0.5	0.525
	1.0			0.475
	0.8	2.45	1.27	0.534
	1.0			0.466

Local approach: one-at-a-time sensitivity measure



Condition variable	Condition levels
Wstat	0,1
Day	0,1,2,3,4,5,6
Gend	0,1
pAge	0,1,2,3,4
SEC	0,1,2,3
Xn-dag	0,1,2,3,4,5

- Step3: Based on our constructed DT, calculate the probability of implement work activity for each input possibility (T=3360).

  E.g., input=(0,0,0,0,0,0)=0.065; input=(1,1,0,0,0,0)=0.679

Global approach: the improved Sobol' method

- Steps of the improved Sobol' method:

  – Step4: The first-order and the total-effect sensitivity indices $\boldsymbol{S_i}$ and $\boldsymbol{S_{T_i}}$ can then be calculated by the following formulas:

$$S_i = \frac{(1/N)\sum_{j=1}^{N} y_B^{(j)}\left(y_{C_i}^{(j)} - y_A^{(j)}\right)}{V(Y)}$$

$$S_{T_i} = \frac{(1/2N)\sum_{j=1}^{N}\left(y_A^{(j)} - y_{C_i}^{(j)}\right)^2}{V(Y)}$$

Where,

$$V(Y) = \frac{1}{N}\sum_{j=1}^{N}\left(y_A^{(j)}\right)^2 - \left(\frac{1}{N}\sum_{j=1}^{N} y_A^{(j)}\right)^2$$

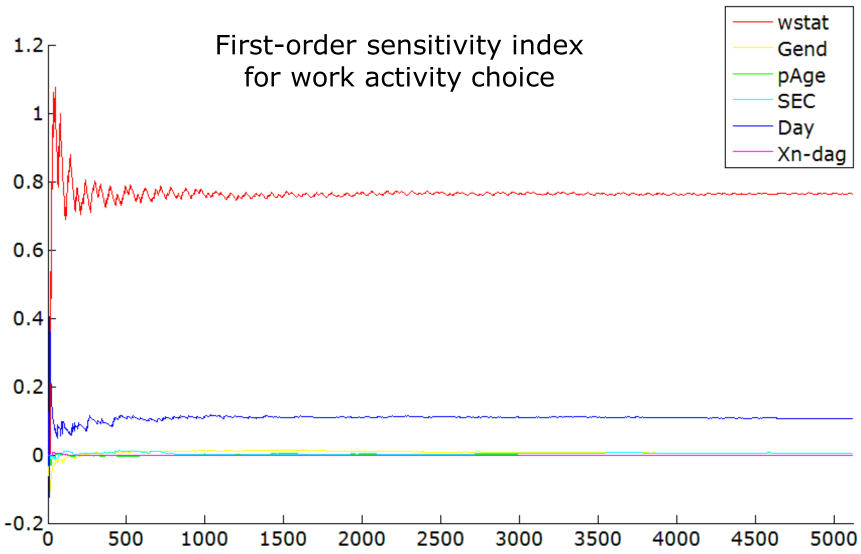
Global approach: the improved Sobol’ method

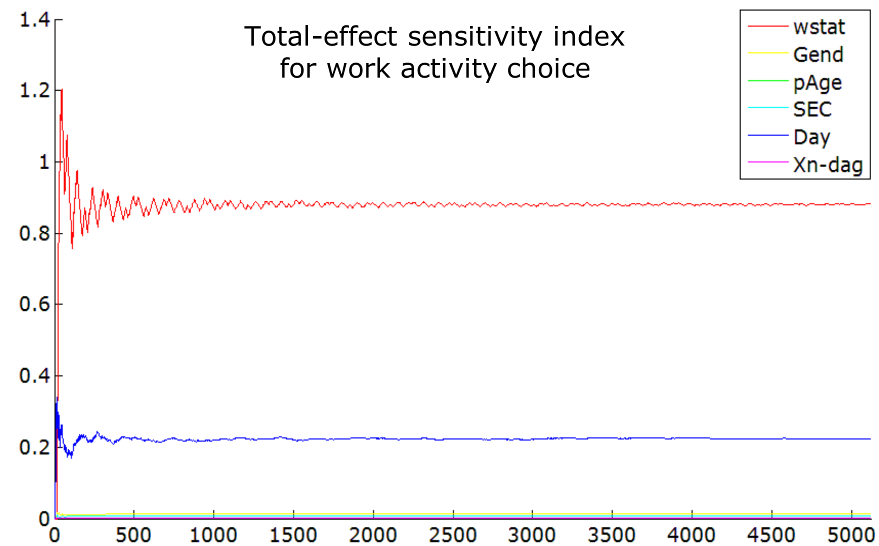
- Steps of the improved Sobol’ method:

  – The most important advantage of applying the above formulas for sensitivity index calculation is the computation is much faster due to existing short cuts, and the total cost would be reduced to N(k+2), which is much lower than the N^2 runs of the original model. For more detailed information on these formulas, we refer to Jansen (1999), Hamm et al. (2006), Saltelli et al. (2006; 2010), and Nossent et al. (2011).

Index	Description
$V(Y)$	The unconditional variance that measures the uncertainty of the model output
S_i	The first-order sensitivity index that quantifies the relative importance of the individual input variable x_i
	A larger value indicates that the input variable is relatively influential on the outcomes of the model
S_{T_i}	The total-effect sensitivity index that accounts for the total contribution to the output variation due to the input variable x_i
	$S_{T_i} = 0$ implies that the input variable can be fixed and is non-influential on the output of the model
$S_{T_i} - S_i$	The value represents the interactions of the input variable x_i with the other input variables
	A larger value means the input variable affects the output mainly through the interactions
$\sum_{i=1}^{k} S_{T_i} > 1 \left(or \ \sum_{i=1}^{k} S_i < 1 \right)$	Interactions exist between the input variables

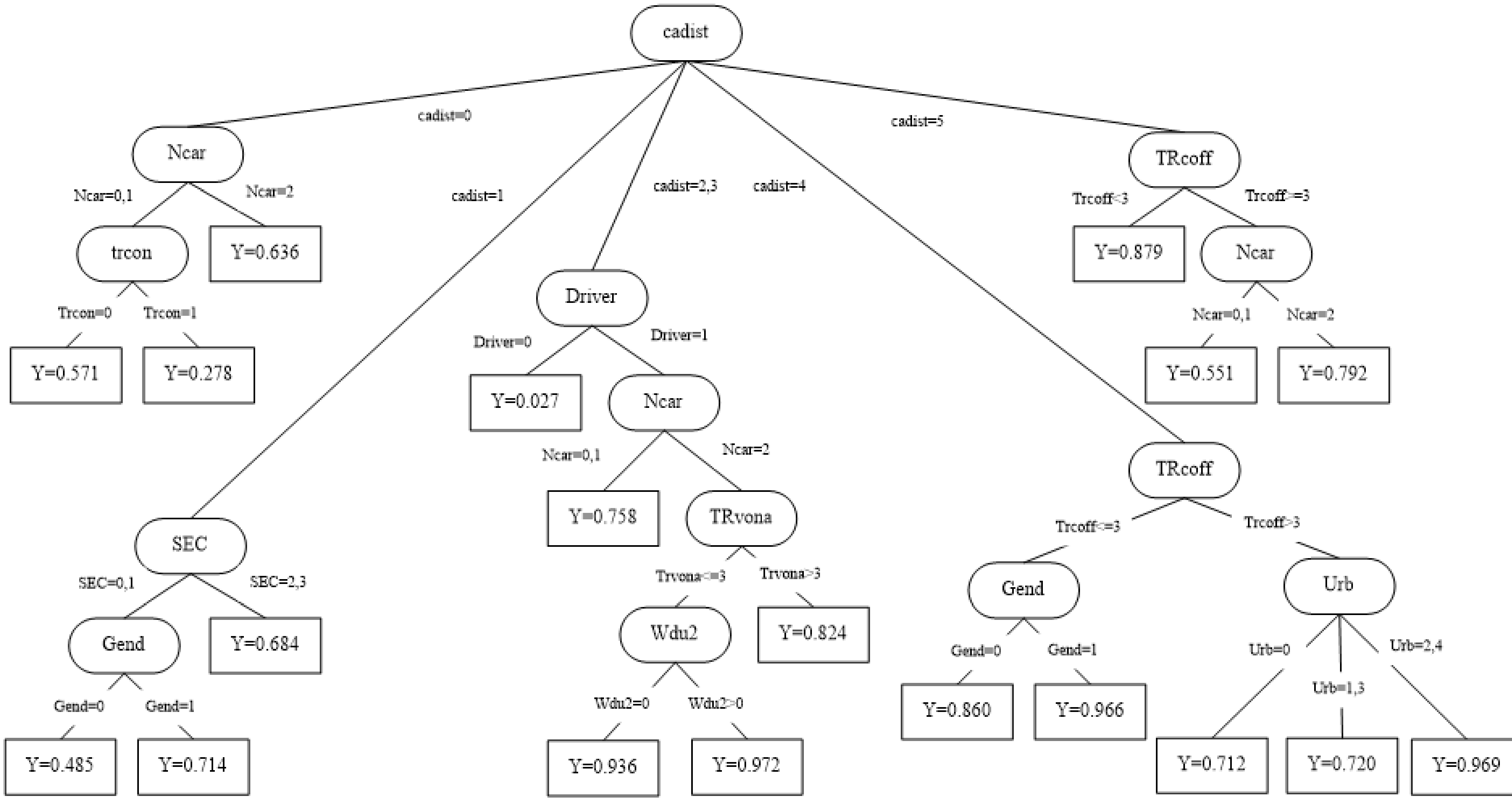
Results of the improved Sobol' method



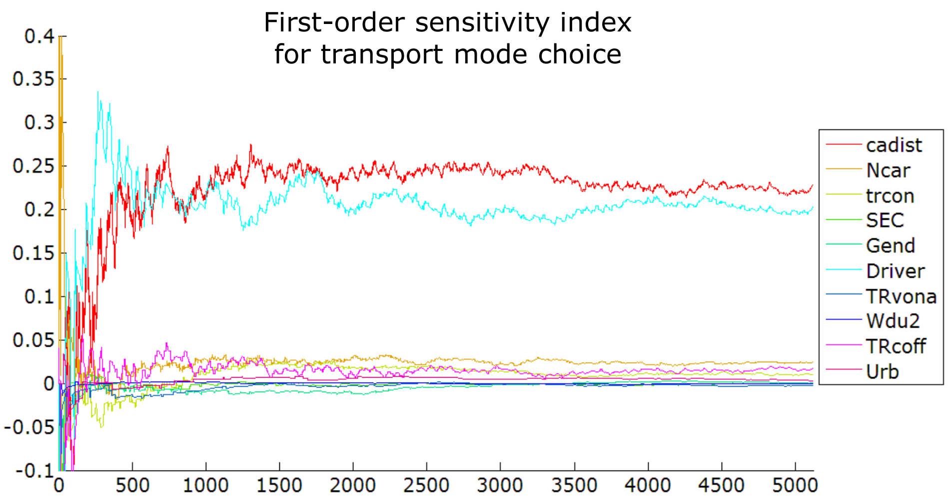


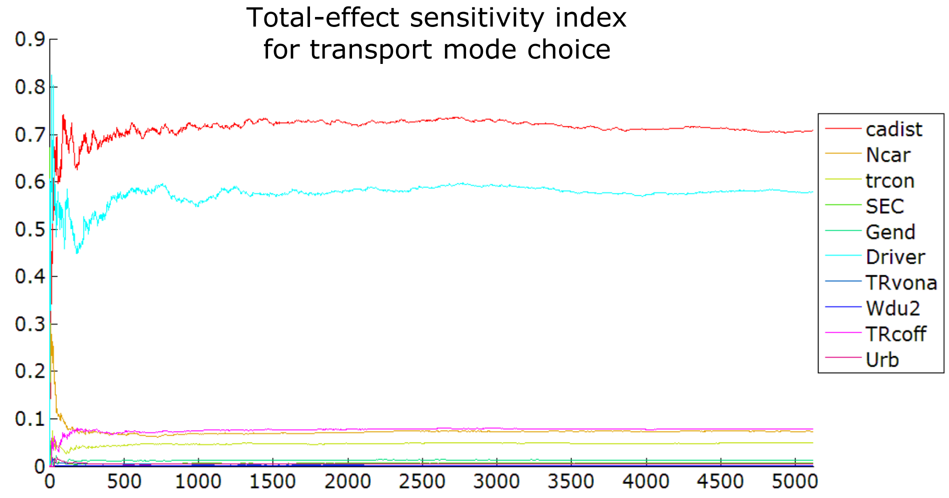
- Results of improved Sobol' method over 5000 times simulation w.r.t. work activity choice decision tree:
  - *Over a certain amount of simulations, both the first-order index and the total-effect index converge to a stable value.*
  - *Both the first-order and the total-effect sensitivity indices for the work activity choice imply: 'wstat' is the most influential variable to the variation of the work choice, followed by 'Day'.*

DT concerning private car mode choice



Results of the improved Sobol’ method





- Results of improved Sobol’ method over 5000 times simulation w.r.t. private car mode choice decision tree:
  - *Over a certain amount of simulations, both the first-order index and the total-effect index converge to a stable value.*
  - *For the transport mode choice, the ‘cadist’ and ‘Driver’ are the two most important variables; the remaining variables have less impacts on the mode choice.*

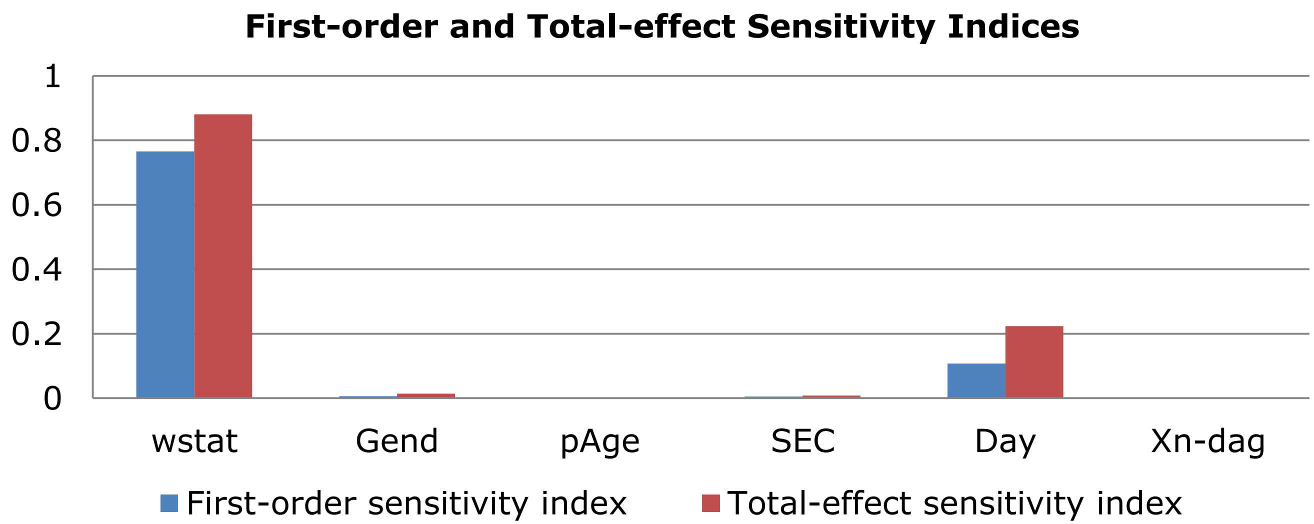
Results of the improved Sobol' method

- Such a result is in line with the one obtained in the OAT sensitivity analysis.
- In order to avoid fluctuation in one simulation, the average values of the last 1,000 simulations are computed for both indices.
  - *The first-order sensitivity index (S_i):*

    quantify the individual impact of input variable
  - *The total-effect sensitivity index (S_{Ti}):*

    account for the total contribution to the choice variation due to the input variable

Results of the improved Sobol' method



GSA results concerning work activity choice-average values of the last 1,000 simulations

S_i : a larger value indicates the input is more influential to the output.

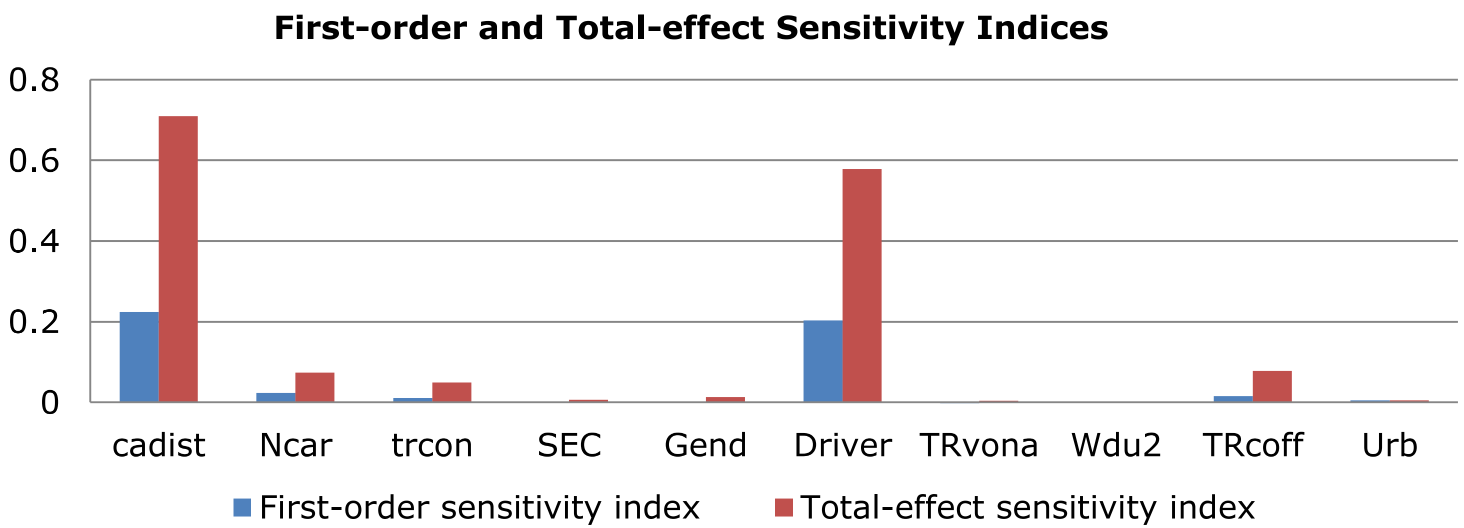
S_{T_i} - S_i : a larger value means the input affect the output mainly through the interactions.

- For work activity choice:

  'wstat' : significant individual effect on the choice variance ($S_i >> S_{T_i} - S_i$).

  'Day': has similar individual and collective effects on the output ($S_i \approx S_{T_i} - S_i$).

Results of the improved Sobol' method



GSA results concerning private car mode choice- average values of the last 1,000 simulations

S_i : a larger value indicates the input is more influential to the output.

S_{T_i} - S_i : a larger value means the input affect the output mainly through the interactions.

- For private car mode choice:

  input variables affect the output mainly through interactions (S_{T_i} - S_i >> S_i).

Lecture summary

- Uncertainty analysis:

  – to quantify the uncertainty around the mean estimate of one or more outcomes.

- Sensitivity analysis:

  – to identify the most significant parameters in the model and to quantify how the input uncertainty influences the outputs.

  - Local approach
    - one-at-a-time sensitivity analysis measure
  - Global approach
    - the improved Sobol' method

Questions

1. Please discuss the main difference between uncertainty analysis and sensitivity analysis.

2. What's the main (dis)advantage of global sensitivity analysis approach contrast to local sensitivity analysis approach?

3. Be able to calculate and compare the Chi-square value (χ^2) for a given frequency table.

Thanks for your attention!

baoqiong@seu.edu.cn