# Sampling Methods

Unit of Biostatistics and Research Methodology,
Universiti Sains Malaysia.
wnarifin@usm.my

# Outlines

- Sampling
  - Probability Sampling
  - Non-probability Sampling

# Learning Objectives

- Understand concept of sampling
- Understand and able to apply methods of probability sampling
- Able to differentiate between probability and non-probability sampling

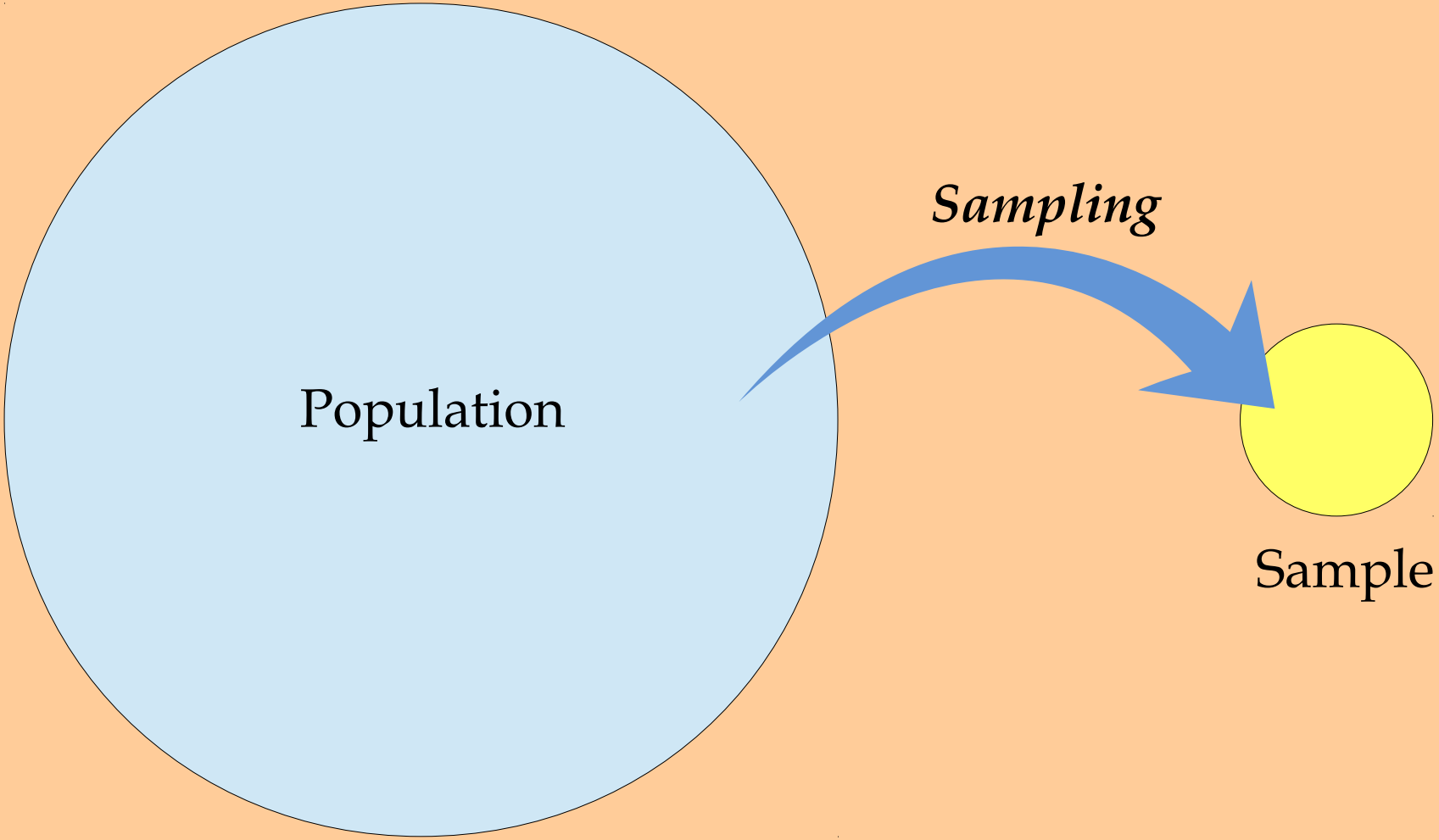
# Sampling

- What is a sample?
- What is sampling?

# Sampling

- Sampling: A process of selecting a number of subjects from a population of interest, so as to make conclusion about the whole population (Everitt and Skrondal, 2010).

# Sampling



# Sampling

- Sub-division (Trochim, 2006):
  - Probability
  - Non-probability

# Probability Sampling

- *Random* selection method.
- Every subject has a probability to be selected (*NOT* necessarily equal probability).
- Probability of selection for each sampling unit is known and > 0.
- It is possible to know how representative a sample of its population.

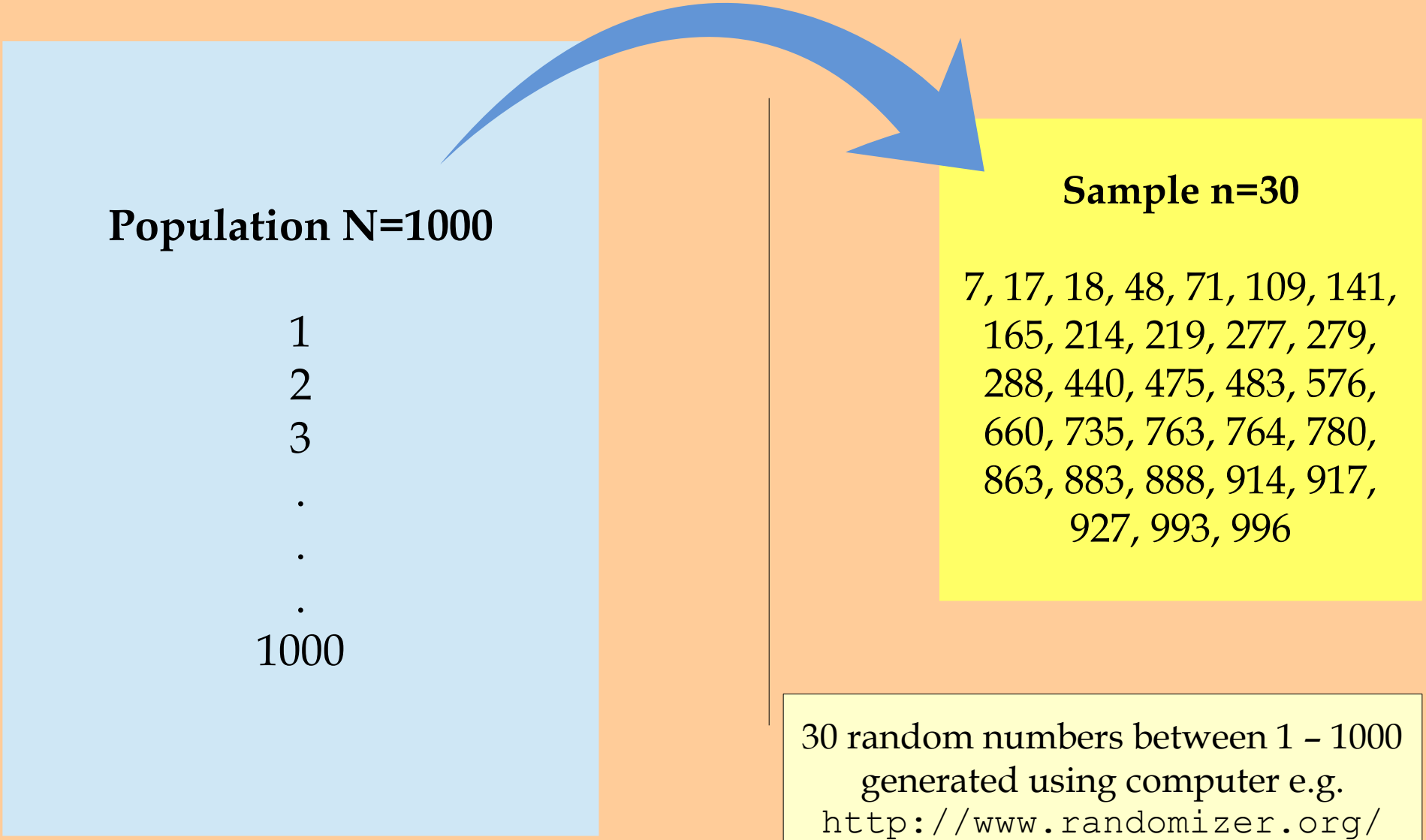
# Probability Sampling

- 4 sampling methods:
  1. Simple random.
  2. Stratified random.
  3. Systematic.
  4. Cluster.

# 1. Simple Random

- A number of distinct subjects are selected randomly from the population in a way that each sampling unit has *equal chance* to be selected.
- Example: 30 patients are randomly selected selected from a list of 1000 patients available to the clinician.

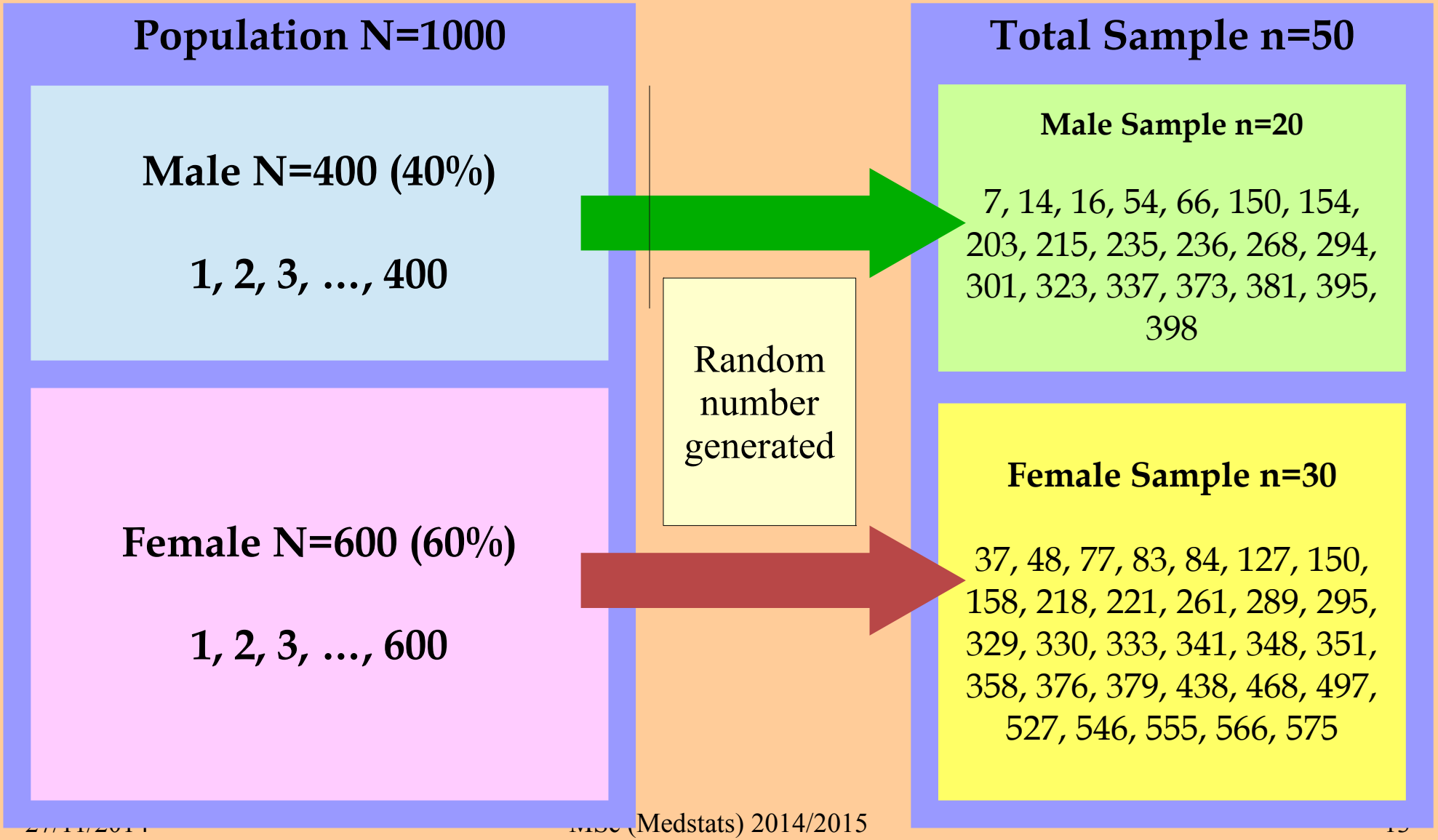
# 1. Simple Random



# 2. Stratified Random

- Subdivision of a population into strata (e.g. gender, race).
- Simple random sampling done within each stratum.
- Ensures major and minority groups are addressed as sampling is done proportionate to strata size in the population.
- Example: For a population consisting of 40% male and 60% female, sampling within each group gives better representation of the population.

# 2. Stratified Random



# 3. Systematic

- Sampling of subjects at a predetermined *sampling interval (k)*, with a random starting number (*j*) in the interval.
- Practical when population list is impossible to obtain (e.g. clinic attendance) but population size (*N*) is estimable (Levy and Lemeshow, 1999).
- Given required sample size (*n*):

$$k = N/n$$

- Starting with *j* subject, followed by every *k*th subject (Trochim, 2006).

# 3. Systematic

- Example: It is estimated that 100 patients come to an oncology clinic per month. To sample 20 patients, more practical to sample starting with a random *j* patient, followed by every *k*th patient.

# 3. Systematic

**1. Interval, $k = N/n$ = 100/20 = 5**

**2. Starting point = Random number between 1 – 5, e.g. 3**

**3. Then every interval of 5**

**Population N=100**

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100

**Sample n=20**

3, 8, 13, 18, 23, 28, 33, 38, 43, 48, 53, 58, 63, 68, 73, 78, 83, 88, 93, 98

# 4. Cluster

- Cluster = Group of people
- Sampling Unit = Cluster e.g. House, Class, Ward etc. → Sampling done on clusters.
- Have to inflate sample size, $n'$ to adjust for cluster effect (Naing, 2011)

$$n' = [1 + (\text{cluster size} - 1)r] \times n$$

- $r$ is correlation between subjects in a cluster → unknown, can assume $r = 0.5$.

# 4. Cluster

**Population N=300**

1, 2, 3, … , 300

**1. 50 houses in area**

**2. On average, 6 persons/house = cluster size**

**Sample n=30**

$n' = [1+(6-1)0.5] \times 30 = 105$

**n of house to sample = 105/6 = 17.5 ≈ 18 houses**

**House, N=50**

**1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50**

**House, n=18**

**2, 7, 11, 12, 15, 18, 20, 22, 25, 26, 30, 31, 32, 33, 36, 38, 43, 44**

# Multistage sampling?

- Any combination of previous 4 sampling methods.

# Activity

- Perform all 4 types of probability sampling methods on this population → *Sampling Methods.xls*

# Non-probability Sampling

- Random selection method not used.
- Selection based on preset criteria set by researcher.
- Could be biased, not representative of population.

# Non-probability Sampling

- Among the methods:
  - Convenient/Accidental/Haphazard:
    - Choose those easily available/sampled
    - e.g. my friends, friends of my friends, relatives, room mates, etc.
  - Purposive:
    - Choose those fulfilling criteria.
    - e.g. only those who come to clinic on Monday, handsome/cute persons only, etc.

# Sampling using SPSS

- Read my article, Arifin (2012).

# References


1. Arifin, W. N. (2012). Random sampling and allocation using SPSS. *Education in Medicine Journal* 4(1), 129-143.
2. Everitt, B., Skrondal, A. (2010). *The Cambridge dictionary of statistics*. 4th ed. New York: Cambridge University Press.
3. Levy, P. S., Lemeshow, S. (1999). *Sampling of populations: Methods and applications*. 3rd ed. New York: John Wiley & Sons.
4. Naing, N. N. (2011). *A practical guide on determination of sample size in health sciences research*. 5th ed. Kelantan, Malaysia: Pustaka Aman Press Sdn. Bhd.
5. Trochim, W. M. K. (2006). *Research methods knowledge base*. Retrieved March 27, 2012, from http://www.socialresearchmethods.net.