MODULE (New)

Non-Conventional Data Sources (Big Data for Gender Equality)

EXERCISES

Curriculum on Gender Statistics Training

This product was developed under the guidance of the Subgroup on Gender Statistics Training, within the Asia-Pacific Network of Statistical Training Institutes.

Exercise 1

Indicate if the following statements are '*True*' or '*False*'.

1. Official statistics can only be produced from data derived from traditional data sources.

2. Administrative records (such as registers) are complementary sources of data for official statistics.

3. The variety characteristic of Big Data refers to the quantity of data generated in near real time.

4. The classification system for Big Data was initially developed by the United Nations Commission for Africa.

5. Human sourced data are collected from sensors and machines used to measure and record events and situations in the physical world.

Exercise 2

1. All of the following are types of Big Data **EXCEPT**:
   a) Structured Data
   b) Semi-Structured Data
   c) Unstructured Data
   d) Macro Data

2. The **THREE** V's that characterize Big Data are:
   a) Volume
   b) Value
   c) Velocity
   d) Variety

3. Which ONE of the following is **NOT** a main challenge to the use of Big Data for gender statistics?
   a) Data processing
   b) Data quality
   c) Privacy
   d) Real time insights

4. Of the data format options listed below, which **ONE** is an example of "unstructured data"?
   a) Scanner data
   b) Bank transaction information
   c) Social media feeds such as posts, comments and shares
   d) Social security numbers

5. Which of the following are opportunities associated with the potential of Big Data for gender statistics:
   a) The potential to easily match data and identify people from whom the data is generated
   b) The availability of larger volumes of data from new sources for more detailed disaggregated statistics

c) The availability of data from a wide range of data sources that can capture diverse perspectives
d) The potential of providing more timely data that can inform policy and decision making

6. Which of the following are **NOT** use cases/ applications of Big Data?
   a) Use of internet searches and social media posts to understand the impact of violence against women
   b) Integration of geospatial data to explore connections between climate change and gender inequality
   c) Integrating multiple indicator cluster data (MICS) with data from Education Management Information System (EMIS) to assess educational outcomes
   d) Use of mobile phone data to estimate population characteristics and monitor migration and urbanization

Exercise 3

Classify each type of data listed below under the three broad categories/ kinds of Big Data sources:

Social Networks (human-sourced information)
Traditional Business Systems (process-mediated)
Internet of Things (machine-generated data)

- e-government data
- Scanner data
- Mobile phone data
- X (formerly known as Twitter) feeds
- Internet searches
- Commercial transactions
- Email or SMS
- Traffic sensors

Solutions to Exercise 1

1. **False**. Official statistics can be produced from both conventional and non-conventional sources.
2. **True.** Administrative records that are compiled through administrative processes such as **vital events recorded through civil registration system is a complementary source of data** and official statistics from censuses and surveys
3. **False;** Variety is one of the three V's characterizing Big Data which refers to the diverse sources of different data types (unstructured, structured, semi-structured)
4. **False;** The classification system was developed by the United Nations Economic Commission for Europe (ECE)
5. **False;** human sourced data are records of human experiences such as social media, email, internet searches

Solutions to Exercise 2

1. d)
2. a), c) and d)
3. d)
4. c)
5. b), c) and d)
6. c)

Solutions to Exercise 3

Traditional Business Systems (process-mediated):

- e-government data
- Commercial transactions
- Mobile phone data

Social Networks (human-sourced information)

- Internet searches
- Email or SMS
- X (formerly known as Twitter) feeds

Internet of Things (machine-generated data)

- Scanner data
- Traffic sensors