

INTERNATIONAL ASSOCIATION FOR OFFICIAL STATISTICS

6-8 DEC. 2016, ABU DHABI

# BIG DATA STRATEGY OF STATISTICS KOREA

HAE RYUN KIM, STATISTICS KOREA

# Outline

- I Overview**.....●
- II KOSTAT Big Data Statistics Strategy**.....●
- III Linked DB using Public-Private big data**.....●
- IV Future Plans**.....●



# **I . Overview**



# Why big data?

- **Providing various information timely for decision making**

- prices, employments, economic development, demographics, etc.\*

\* **Source of big data : Administrative data, Transaction data, Online data, Tracking Device data, Sensor data, etc.**

- **Reducing the cost for producing statistics**

- easily obtaining data sources from internet, financial transaction, etc. without traditional survey

- **Creating future growth engines**

- IoT, Big data technology, Analytics, etc.

# Restriction on the use of Big Data for official statistics

## □ **Data Access**

- Most of big data are from private sector which can be used for their business
- No obligation to provide data for statistical purpose

## □ **Privacy Issue**

- Data are valued when using private information but strict privacy law

## □ **Quality problem**

- Weak representativeness, unstable data collection, etc.



## **II . KOSTAT Big Data Statistics Strategy**



# Progress in KOSTAT

- **Implementing pilot study using private big data (2013-2014)**
  - (Online Daily Price Index) In 2013, Price Index using 284 price data from online shopping mall
  - (Daily Population Movement Analysis) In 2014, using SK Telecom data to analyze daily population movement in 3 areas: Busan, Ulsan and Gyeongnam
- **Establishing new division for Big data (2015. 10.1) : Big Data Strategy, Production, System Team**

# Big data Strategy

- ◆ Produce various statistical information by linking **public data** with **private sector big data**
- ◆ Establish **framework** & Support **various activities** on the use of big data

**1** Link/analyze public data and private sector big data

**2** Support policies using private sector big data

**3** Support private sector big data activities

**4** Establish grounds for the use of big data

**5** Establish external cooperation system



# **(1) Link/analyze public data and private sector big data**

## **□ Link KOSTAT's public data\* with private sector big data for analysis**

- Practicing various case study under the privacy protection law framework

**\* 93 administrative data and 40 survey data**

Established KOSTAT's own linkage methodology and produced "Newly married couples statistics" in accordance with

**Statistics Law & Privacy Protection Law**

## **(2) Support policies using private sector big data**

- **Develop indicator system on economic and social situation**
  - Indicators on people's livelihood\* in timely manner (weekly, monthly)
    - \* 9 indicators on price, employment, household finances, self-employed finances
  
- **Improve daily online price index**
  - \* 6 online shopping malls and 284 items
  
- **Producing consumer sentiment index**
  - Using text data collected from blogs, SNS, and other types of internet posts

## **(3) Support private sector big data activities**

### **□ Validation of forecasts**

- Validate accuracy and reliability of forecast results published by private sector using big data

### **□ Masking assistance and consulting**

- Provide masking assistance and consulting on techniques to provide governmental big data to private sector

## (4) Establish grounds for the use of big data

### □ **Data sharing system**

- Establish "Statistical Data Center" for sharing public data and provide data linkage service

\* **De-identified & Standardized data, Linkage Public-Private data, Integrated DB**

### □ **Institutional improvements**

- Prepare legal framework to access and to use private sector big data

### □ **Methodological research**

- Study methodologies on the limitation of big data

\* **Weak representation, unable to produce estimations due to absence of statistical population, unstable data collection, etc.**

# (5) Establish external cooperation system

## □ **Within Korea**

- Organize “Big-data Strategy Forum” to discuss issue on the use of big data
- Obtain and exchange data between relevant agencies through MOUs\*

\* **NAVER(2015. 9.), KCB(2016. 1.), Lotte Point(2016. 9.),  
Korea Federation of Credit Guarantee Foundation(2016. 9.),**

## □ **Internationally**

- Join UN’s GWG on Big Data and develop network with international experts
- Cooperation with CBS

# Cooperation with CBS

## □ **Institutional cooperation**

- MOU on big data during Korea-Netherlands Summits ('16. 9. 26.)
- (Area) Obtaining data, Developing techniques, Data linkage & privacy, Methodology, E-learning, Staff exchange

## □ **Benefits from cooperation**

- Mutually learning know-how in the area of comparative advantage\* of each country

\* (KOSTAT) big data linkage, access of private data, privacy issue  
(CBS) SNS, censor data



**III. Case on Linked DB using  
Public-Private Big Data :  
DB on newly married couples' debt**



# Background

- **The recent increase of political interest in household debt demands exact household debt information especially on micro-level**
    - There is no debt statistics by various household characteristics such as single-, old aged-, self-employed-household, etc.
    - \* **Macro-level : Household credits, Bank of Korea/ Micro-level : Household finance & welfare statistics, KOSTAT(20,000 h/h)**
- ⇒ Newly married couples are of importance for housing policy and low fertility rate policy



# Used Big Data : Public & Private data

- **MOU between KOSTAT & Private sector\***
  - Data access and joint-research
    - \* **Individual level debt**
- **KOSTAT's Survey & Administrative data**
  - Population movement data, Census data, Owing House statistics, National Insurance data
- **Private data from Credit evaluation company**
  - Debt, Grade of credit, Delinquency, card spending, etc.

# Methodology of Linkage

- **Step 1 : Data sharing & de-identification**
  - KOSTAT provide de-identified data to private company according to Statistics Law (Para. 31)
  
- **Step 2 : Data linkage**
  - Private company link both data-set using linkage key(anonymized identifier) and delete linkage key
  
- **Step 3 : Post Management**
  - All process done at KOSTAT's data center (no intranet & internet)
  - Only aggregated data can be carried out outside

# Definition & Analysis

## □ **Cross-sectional data**

- Newly married couples being married for 5 years or less since its report in 2014
- Sampling 50,000 from 1,540 thousand couples

## □ **Longitudinal data**

- Newly married couples in 2014
- Sampling 50,000 from 294 thousand couples

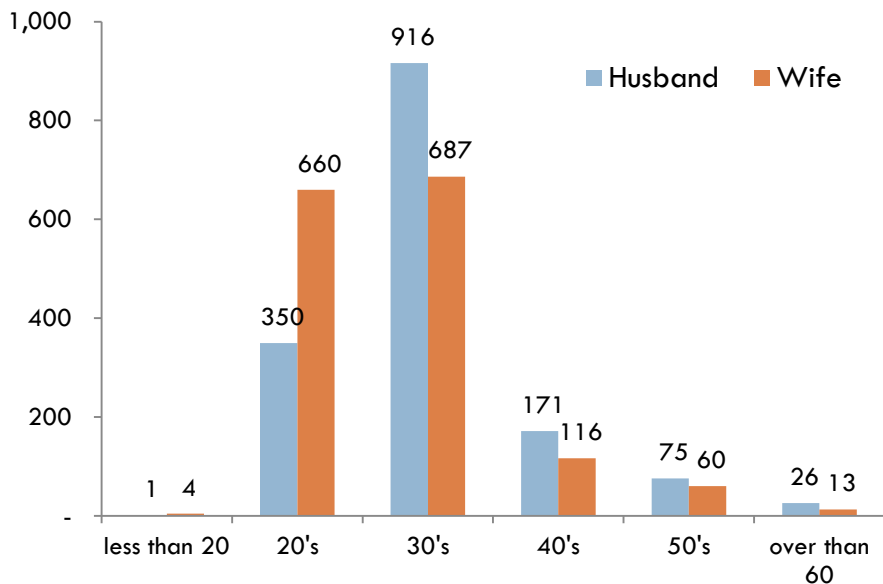
## □ **Analysis**

- Income, Debt, No. of child, No. of House, Card consumption, etc.

# Analysis Results

## No. of New Married couple

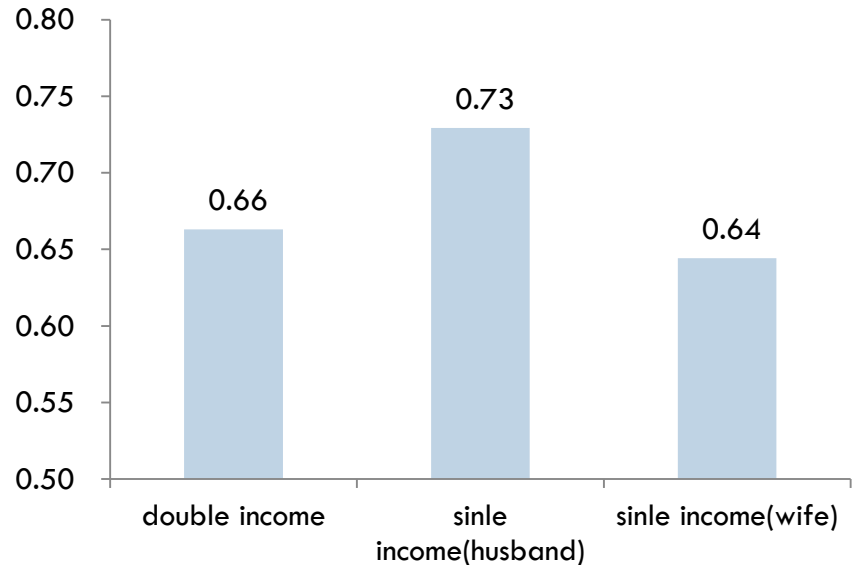
(Unit: 1,000 person)



- Total 1,539 thou. couples (1.7%)  
- 46% in Seoul & Gyeonggi-province
- Husband & wife are both 30's (36.0%)

## No. of Child by Economic activity

(Unit: person)

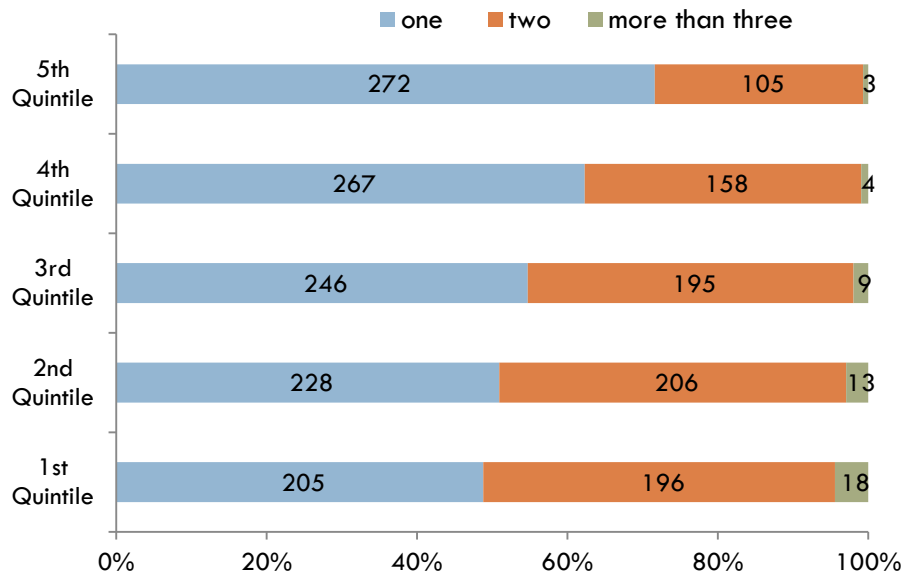


- Avg. no. of child is 0.68  
- Economic activity of wife is important for the birth rate

# Analysis Results

## No. of Child by Income Group(Longitudinal DB)

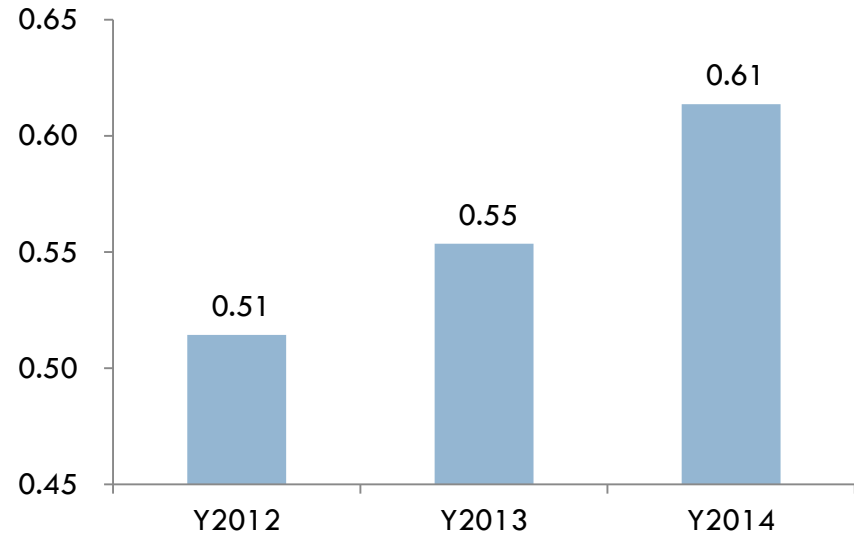
(Unit : 100 households)



- Y2014 : 1<sup>st</sup> Q(1.1 child), 5<sup>th</sup> Q(0.8 child)
  - The no. of child is lower in higher income group

## No. of House(Longitudinal DB)

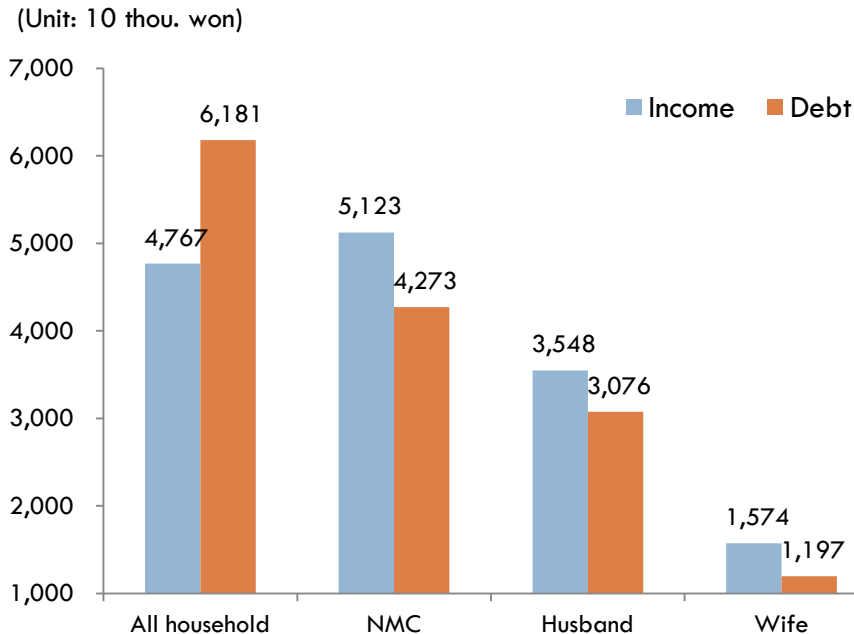
(Unit : house)



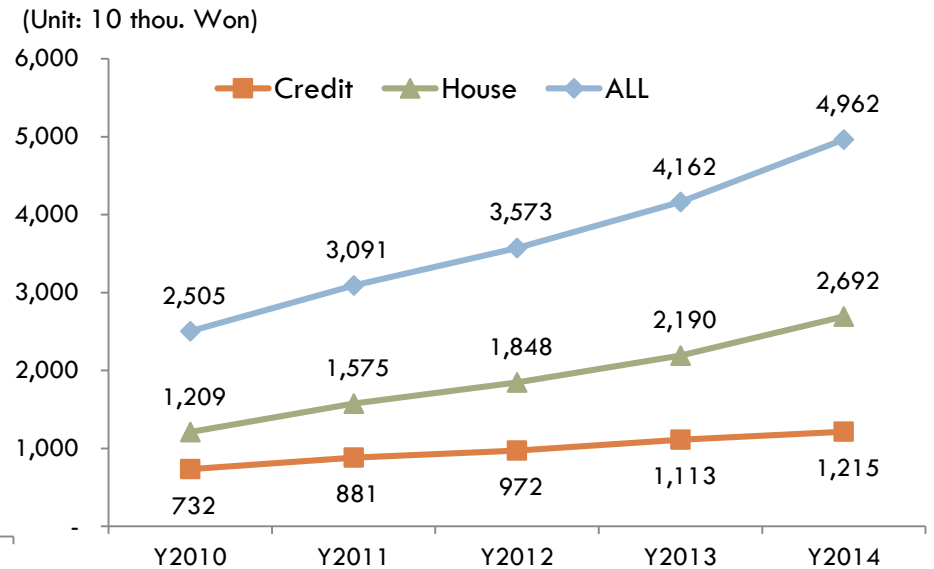
- Still newly married couple have higher demand on houses

# Analysis Results

## Income & Debt



## Debt by Types of Loan(Longitudinal DB)



- (Y2014) Avg. Debt 4,273 10 thou. won, Avg. Income 5,123 10 thou. won  
- Income/Debt as 0.8 is lower than all household 1.3

- Housing mortgage loan increases rapidly



## **IV. Future Plans**



# Future Plans

- **Expand linked DB on Household Debt**
  - single person household, old aged household, etc.
  - linking 20% sample dataset of Population census
- **Build other linked DB**
  - Data from credit guarantee for small business, from Mobile network operator, from Credit card, etc.
- **Establish Statistical Data Center**
  - Information Strategy Plan



**Thank you!!**