Real-Time Analysis of 2020 Census Internet Self-Response Paradata

An Overview of Census Real-Time Collection of Web User Activities through Internet Self-Response Operation

FedCASIC 2020

Lydia Shia, U.S. Census Bureau

Decennial Statistical Studies Division

Disclaimer: Any views expressed are the author's and not those of the U.S. Census Bureau

Overview

- Introduction to paradata's Role in the Internet Self-Response Operation
- Paradata Data Structure
- Real-Time Analysis of Internet Response Data
  - Assessing and monitoring data quality in real-time for operation management
  - Explorations of paradata through summary statistics
- Limitations and Next Steps

United States® Census Bureau

An Introduction to Paradata's Role in Internet Self Response (ISR)

The 2020 U.S. Decennial Census was the country's first census where most users responded via the internet. In fact, approximately 80% of those who self-responded used the ISR. The anticipated benefits of using ISR were:

- Increased flexibility
- Expanded ability to answer in different languages
- Reduced costs for materials
- Reduced costs of field work
- Reduced the risk of in-person contact during the 2020 pandemic

To ensure data quality and instrument performance, the instrument collects paradata that contains a variety of factors reflecting respondents' navigation behaviors such as login, breakoff and completion time.

United States® Census Bureau

Real-Time Analysis of Internet Response Data

Response rate in relation to mailing reminders

Panel	# of Cohorts	Mailing 1	Mailing 2	Mailing 3*	Mailing 4*	Mailing 5*	Mailing 6*	Mailing 7*
Internet First	4	Letter	Letter	Postcard	Letter + Questionnaire	"It's not too late" Postcard	Postcard	Letter + Questionnaire
Internet Choice	N/A	Letter + Questionnaire	Letter	Postcard	Letter + Questionnaire	"It's not too late" Postcard	Postcard	N/A
***= Targeted only to non-respondents**								

United States® Census Bureau

Paradata Data Structure

A simple overview

Three key variables that paradata captures

Session ID

- Key that represents individual sessions within a household.

Name

- Label that describes screen and instrument activities.

Value

- Field that captures user data such as timestamp, IP, screen language, device info, etc.

United States® Census Bureau

Paradata Data Structure

Screen duration calculation on mock paradata

Session ID	Name	Value
Session1	PD_SCREEN_A_START_TIME	2020-03-09T23:26:39:502Z
Session1	PD_SCREEN_A_CLOSE_BROWSER_TIME	2020-03-09T23:26:43:326Z
Session2	PD_SCREEN_G_LOAD_TIME	2020-03-10T08:30:05:924Z
Session2	PD_SCREEN_J_LOAD_TIME	2020-03-10T08:30:05:924Z
Session2	PD_SCREEN_O_LOAD_TIME	2020-03-10T08:30:08:769Z
Session2	PD_SCREEN_G_CLOSE_BROWSER_TIME	2020-03-10T08:30:08:769Z

United States® Census Bureau

Paradata Data Structure

Screen duration calculation on mock paradata

Session ID	Name	Value
Session1	PD_SCREEN_A_START_TIME	2020-03-09T23:26:39:502Z
Session1	PD_SCREEN_A_CLOSE_BROWSER_TIME	2020-03-09T23:26:43:326Z
Session2	PD_SCREEN_G_LOAD_TIME	2020-03-10T08:30:05:924Z
Session2	PD_SCREEN_J_LOAD_TIME	2020-03-10T08:30:05:924Z
Session2	PD_SCREEN_O_LOAD_TIME	2020-03-10T08:30:08:769Z
Session2	PD_SCREEN_G_CLOSE_BROWSER_TIME	2020-03-10T08:30:08:769Z

United States® Census Bureau

Paradata Data Structure

Screen duration calculation on mock paradata

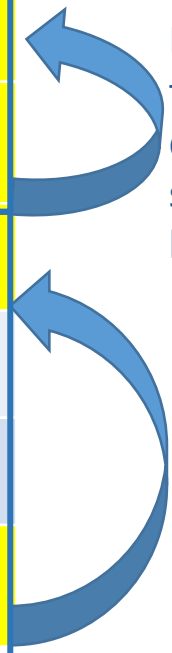
Session ID	Name	Value
Session1	PD_SCREEN_A_START_TIME	2020-03-09T23:26:39:502Z
Session1	PD_SCREEN_A_CLOSE_BROWSER_TIME	2020-03-09T23:26:43:326Z
Session2	PD_SCREEN_G_LOAD_TIME	2020-03-10T08:30:05:924Z
Session2	PD_SCREEN_J_LOAD_TIME	2020-03-10T08:30:05:924Z
Session2	PD_SCREEN_O_LOAD_TIME	2020-03-10T08:30:08:769Z
Session2	PD_SCREEN_G_CLOSE_BROWSER_TIME	2020-03-10T08:30:08:769Z

United States® Census Bureau

Paradata Data Structure

Screen duration calculation on mock paradata

Session ID	Name	Value
Session1	PD_SCREEN_A_START_TIME	2020-03-09T23:26:39:502Z
Session1	PD_SCREEN_A_CLOSE_BROWSER_TIME	2020-03-09T23:26:43:326Z
Session2	PD_SCREEN_G_LOAD_TIME	2020-03-10T08:30:05:924Z
Session2	PD_SCREEN_J_LOAD_TIME	2020-03-10T08:30:05:924Z
Session2	PD_SCREEN_O_LOAD_TIME	2020-03-10T08:30:08:769Z
Session2	PD_SCREEN_G_CLOSE_BROWSER_TIME	2020-03-10T08:30:08:769Z



End time – Start time = time duration of session id1 on landing screen

End time – Start time = time duration of session id2 on dashboard screen

United States® Census Bureau

Real-Time Analysis of Internet Response Data

An overview of metrics on paradata dashboard for real-time analysis

1. Overall device type distribution within session completion status
2. Time spent in the instrument
3. Number of daily logins in relation to mailing dates
4. Weekly breakoff rate
   - By device and session type
   - By household size
   - By geography

United States® Census Bureau

Real-Time Analysis of Internet Response Data

What we were able to accomplish through real-time analysis on paradata:

- Monitored number of submits and instrument time to study user behaviors triggered by contact strategy
- Studied the adoption of new technologies such as device type, and their effect on survey completion and breakoff
- Extracted response patterns through language selection
- Monitored breakoff rate by household size and ID type
  - Hispanic screen
  - College campus vacancy
- Investigated data reliability through IP analysis

United States® Census Bureau

Limitations and Next Steps

- Limitations
    - Occasional inaccurate recording of paradata
    - Occasional missing paradata
    - Long system processing times due to large scale datasets
    - Inability to differentiate a breakoff between involuntary vs voluntary breakoff status; in other words, we do not know if the user closes out the browser vs the browser crashed

- Next Steps
    - 2020 ISR assessment
    - Using data captured to further research on Census user experiment studies

United States® Census Bureau

Contact Info

Lydia Shia

Lydia.shia@census.gov

United States® Census Bureau

References

- ISR Operation Memo - https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/ISR_detailed_operational_plan.pdf

United States® Census Bureau