Characteristics of households using electronic questionnaire in the 2011 Population Census in Hong Kong

T.S. Tsang
Census and Statistics Department, Hong Kong, China
tstsang@censtatd.gov.hk

Abstract

A multi-model data collection approach was first adopted in the 2011 Population Census (11C) in Hong Kong where respondents were allowed to provide the required information by self-enumeration in addition to the traditional mode of face-to-face interview by enumerators. In the 11C, electronic questionnaire (e-Q) was available for both long form and short form questionnaires for self-enumeration whereas questionnaire in paper form for postal return was only available for short form questionnaire. About 13% of the households had used e-Q for data reporting. This paper aims to study the characteristics of these households. The logistic regression model technique is used to identify the major factors affecting the e-Q usage and examine the substitution effects between the paper mode and the e-Q mode for households of different characteristics.

Keywords: Data collection mode, electronic questionnaire, population census, Internet penetration, logistic regression model

1. Introduction

It is an established practice since 1961 for Hong Kong to conduct a population census every ten years and a by-census in the middle of the intercensal period. The 2011 Population Census (11C) was conducted in June to August 2011. A multi-modal data collection approach was adopted in the 11C.

The whole 34-day data collection period was divided into two phases. In the first phase, households[1] were asked to provide information by self-enumeration including submission of electronic questionnaires (e-Q) and postal questionnaire. The e-Q mode was available to households selected for both short form and long form enumeration[2] whereas postal questionnaires were provided to households selected for short form enumeration only. In the second phase, face-to-face interviews were conducted by enumerators for households with no return during the first phase.

Overall speaking, the e-Q take-up rate (defined as the number of households using e-Q as the reporting mode over all households) in the 11C was 13%. Analysed by form type, the e-Q take-up rate for long form households was 35% and that for short form households was lower at 10% only. Another 48% of short form households opted to provide the required information by mailing back the completed postal questionnaire possibly because of its simple design (with all questions contained in one A3-sized paper) and availability of pre-paid return envelope.

---

[1] Notification letters were issued to households living in quarters with valid postal addresses. However, self-enumeration modes were not available to households living in temporary structures in segment areas with incomplete addresses and households living in quarters split from an existing unit. As there were very few households unable to use self-enumeration modes, they were not singled out for the analysis in this paper.

[2] There were 11 data topics in the short form questionnaire and 41 data topics in the long form questionnaire.

The e-Q data collection in the 11C was an operational success. More than 95% of households registering the use of e-Q had eventually completed and submitted their e-Q over the Internet. There were very few issues on data quality and almost all e-Q households were willing to provide supplementary information during the follow-up by phone. The e-Q mode is very cost effective and is therefore highly recommended for adoption in future rounds of population census/by-census. How to boost the e-Q take-up rate will be an important area to be studied.

This paper aims to study the characteristics of households using e-Q with a view to facilitating the planning of future rounds of population census/by-census. In the first part of this paper, a logistic regression model is used to identify the major factors affecting the e-Q usage in the 11C. In the second part, another logistic regression model is used to examine the substitution effects between paper questionnaire and e-Q modes for households of different characteristics.

2. Characteristics of households using e-Q

Household data from the long form dataset are used to study the characteristics of households using e-Q (e-Q households). For the sake of comparison, the characteristics of the other enumerated households (non e-Q households) are also presented. Figures presented in this section are compiled based on unweighted long form data (with non-response cases excluded) and cover domestic households[3] only.

Type of housing

38% of e-Q households were owner occupiers living in private residential flats. This proportion was substantially higher than that of 27% among non e-Q households. Another 18% of e-Q households were living in subsidized home ownership housing and this proportion was also higher than that of 13% among non e-Q households.

Age composition of household members

28% of e-Q households had at least one youth member aged 15-24 and this proportion was higher than that of 23% among non e-Q households. 47% of e-Q households had at least one prime adult member aged 25-44 whereas the proportion was 42% among non e-Q households. The proportions of households with at least one middle-aged member aged 45-64 among e-Q households and non e-Q households were 45% and 36% respectively. Analysed by age of household head, there were relatively higher proportions with household head aged 25-44 and 45-64 among e-Q households (32% and 53% respectively) than non e-Q households (25% and 48% respectively).

Sex of household head

62% of e-Q households were headed by a male member. This proportion was similar to that of 65% among non e-Q households.

Ethnicity of household head

98% of heads of e-Q households were of Chinese ethnicity, comparing to that of 96% among non e-Q households. The proportions of households headed by a person of Asian ethnicity (other than Chinese) and a person of non-Asian ethnicity among e-Q households were 0.5% and 1.7% respectively. The corresponding proportions among non e-Q households were 2.5% and 1.4% respectively.

---

[3] Excluding collective households and households with all members being mobile residents as defined under the Hong Kong Resident Population approach.

Economic activity status of members

Only 12% of e-Q households had no members in the labour force. This proportion was lower than that of 23% among non e-Q households. On the other hand, about 38% of e-Q households had all adult members (i.e. those aged 18 and over) in the labour force. This proportion was slightly higher than that of 35% among non e-Q households.

Highest educational attainment (among all household members)

About 41% of e-Q households had at least one member who had attended education in degree course. This proportion was much higher than that of 27% among non e-Q households. On the other hand, only 5% of e-Q households had only members attaining primary (or below) education level. This proportion was lower than that of 13% among non e-Q households.

Highest occupation group (among all household members)

There were higher proportions with at least one of the household members working as managers and administrators, professionals, associate professionals or clerical support workers among e-Q households (66% collectively) than non e-Q households (44%).

3. Factors affecting the use of e-Q

12 variables with reference to the characteristics mentioned in the previous section are used to fit a logistic regression model. For simplicity, this model is referred to as Model 1 hereafter. All the 12 variables are found statistically significant at the 1% significance level (see Table 1).

The odds ratios of Model 1 are presented in Table 2. Some salient observations are as follows:

- Type of housing: Households living in staff quarters, those in subsidized home-ownership housing and owner-occupiers in private residential flats were more likely to use e-Q. The odds ratios relative to households living in non-domestic and temporary housing were generally very high since many of the households living in non-domestic and temporary housing did not receive the notification letter for registering the use of e-Q.
- Age composition: Households headed by a prime adult member were more likely to use e-Q but the odds would decrease if there was more than one prime adult member in the household.
- Sex of household head: Households headed by a female member were more likely to use e-Q than those headed by a male member.
- Ethnicity of household head: Households headed by a person of Asian ethnicity (other than Chinese) were less likely to use e-Q. The odds were only about one-fifth for those headed by a person of non-Asian ethnicity.
- Economic activity status: Households with one adult member in the labour force and another adult member not in the labour force were more likely to use e-Q. On the other hand, the odds of households with all adult members in the labour force were even lower than the odds of households with no members in the labour force.
- Highest educational attainment: Households with at least one of the members attaining post-secondary education (either in degree course or non-degree course) were more likely to use e-Q.
- Highest occupation group: The odds of households using e-Q would increase if there were members engaged as managers and administrators or professionals.

4. Substitution effects between paper questionnaire mode and e-Q mode

Data in both long form and short form datasets are used to study the substitution effects. It is assumed that short form households of the same characteristics as long

form households will use e-Q if no postal questionnaire is provided.

Only the three variables relating to the characteristics of household head (namely age, sex and ethnicity of household head) are used for model fitting. A dummy variable (i.e. whether the household was selected for short form enumeration) is added in the logistic model. To examine the substitution effects, the interaction terms are added in the model. For simplicity, this model is referred to as Model 2 hereafter.

All the variables in Model 2 are found statistically significant at the 1% significance level (see Table 3). They can be used to explain the mode preferences (between paper questionnaire and e-Q) among different groups of respondents. The results of Model 2 (see Table 4) indicate the following:

- Households with older household head were more likely to switch to use postal questionnaire if it was offered. The odds ratio between the two scenarios with and without postal questionnaire mode was 0.26 for households with head of household aged 45-64.
- Households headed by a female member were more likely to switch to use postal questionnaire if offered. The odds ratio for households headed by a female member was 0.25.
- Households headed by a person of non-Asian ethnicity or those headed by a person of Chinese ethnicity were more likely to switch to use postal questionnaire if offered.

5. Conclusion

The two different logistic regression models in this paper provide insight on the e-Q usage in the 11C from different perspectives. For example, it is noted that households headed by a middle-aged person and those headed by a person of non-Asian ethnicity were very willing to use e-Q but they would switch to use postal questionnaire if offered. These results can be used as useful reference for the planning of the forthcoming 2016 Population By-census.

References

Census and Statistics Department (2012a) *2011 Population Census Main Report: Volume I*. Hong Kong: Census and Statistics Department

Census and Statistics Department (2012b) *2011 Population Census Main Report: Volume II*. Hong Kong: Census and Statistics Department

Lee E., Kim S. (2012) *Factors affecting the decision to participate in the internet option for the 2010 Census of Korea*. Paper presented at the United Nations International Seminar on Population and Housing Censuses: Beyond the 2010 Round, Seoul 27-29 November 2012.

Valente P. (2012) *Use of the Internet response for censuses of the 2010 round in the UNECE region*. Paper presented at the United Nations International Seminar on Population and Housing Censuses: Beyond the 2010 Round, Seoul 27-29 November 2012.

Table 1: Analysis of effects of Model 1

Variable	Degree of freedom	Wald Chi-square	P-value
Type of housing	6	2 799	<0.0001
No. of children (aged under 15)	2	82	<0.0001
No. of youths (aged 15-24)	2	18	<0.0001
No. of prime adults (aged 25-44)	2	182	<0.0001
No. of middle-aged (aged 45-64)	2	305	<0.0001
No. of older persons (aged 65+)	2	137	<0.0001
Age of household head	3	1 399	<0.0001
Ethnicity of household head	2	856	<0.0001
Sex of household head	1	291	<0.0001
Economic activity status	2	190	<0.0001
Highest educational attainment	4	2 117	<0.0001
Highest occupation group	6	1 811	<0.0001

Table 2: Odd ratio estimates of Model 1

Variable		Point Estimate	95% Wald Confidence Limits	
Type of housing *(versus Non-domestic & temporary housing)*	Public rental housing	15.36	12.84	18.37
	Subsidized home ownership housing	19.39	16.20	23.21
	Private residential flats – Owner occupiers	17.59	14.70	21.04
	Private residential flats – Non owner occupiers	11.16	9.32	13.37
	Staff quarters	22.20	18.22	27.06
	Other quarters in private permanent housing	9.03	7.53	10.84
No. of children *(versus 2 members)*	0 member	1.18	1.14	1.22
	1 member	1.13	1.09	1.18
No. of youths *(versus 2 members)*	0 member	1.06	1.03	1.10
	1 member	1.09	1.05	1.13
No. of prime adults *(versus 2 members)*	0 member	1.30	1.25	1.36
	1 member	1.22	1.17	1.26
No. of middle-aged *(versus 2 members)*	0 member	0.62	0.59	0.66
	1 member	0.73	0.68	0.77
No. of older persons *(versus 2 members)*	0 member	0.63	0.58	0.69
	1 member	0.69	0.64	0.76
Age of household head *(versus 65+)*	< 25	1.57	1.42	1.73
	25-44	1.96	1.90	2.04
	45-64	1.53	1.49	1.58
Sex of household head *(versus Female)*	Male	0.84	0.82	0.86
Ethnicity of household head *(versus Others)*	Chinese	0.94	0.87	1.01
	Asian (other than Chinese)	0.20	0.18	0.23
Economic activity status *(versus all adult members in labour force)*	No members in labour force	1.07	1.02	1.11
	With one member in labour force and another adult member not in labour force	1.19	1.16	1.22

Variable		Point Estimate	95% Wald Confidence Limits	
Highest educational attainment *(versus Post-secondary: degree course)*	Primary and below	0.48	0.45	0.50
	Lower secondary	0.48	0.46	0.50
	Upper secondary/Sixth form	0.75	0.73	0.77
	Post-secondary: Non-degree	1.11	1.08	1.14
Highest occupation group *(versus Others)*	Managers and administrators	1.86	1.79	1.95
	Professionals	1.76	1.67	1.85
	Associate professionals	1.36	1.31	1.41
	Clerical support workers	1.45	1.40	1.51
	Service and sales workers	0.89	0.86	0.93
	Elementary occupations	0.88	0.84	0.92

Table 3: Analysis of effects of Model 2

Variable	Degree of freedom	Wald Chi-square	P-value
Age of household head	3	5 676	<.0001
Sex of household head	2	2 025	<.0001
Ethnicity of household head	1	18	<.0001
Short form	1	469	<.0001
Age of household head x Short form	3	843	<.0001
Ethnicity of household head x Short form	1	142	<.0001
Sex of household head x Short form	2	77	<.0001

Table 4: Odd ratios of Model 2 (with postal questionnaire mode versus without postal questionnaire mode)

Variable		Point Estimate
Age of household head	< 25	0.49
	25-44	0.35
	45-64	0.26
	65+	0.31
Sex of household head	Male	0.28
	Female	0.25
Ethnicity of household head	Chinese	0.18
	Asian (other than Chinese)	0.26
	Others	0.14