Canadian Internet Use Survey Analysis

Background & Objective

The Canadian Internet Use Survey has been conducted every 2 years to learn about household and individual internet usage. It is a revamped version of the Household Internet Use Survey which only focused on the general household and not individuals within the household. This version seeks to learn about individual differences in internet use, using a Household component and an Individual Component. Household component focuses on connection speed, access to internet and household income while the individual component focuses on specific use, electronic commerce, privacy and security experiences the user goes through while using the internet. The objective of this survey is to obtain reliable estimates of internet access in rural areas. It also is used to obtain information on the types of goods people purchase over the internet, online behaviour, intensity of internet use, views on online privacy and security as well as the reasons some members do not use the internet.

Survey Methodology

At the time of the survey collection, Oct 14th, 2012 – Nov 20th, 2012, the sampling frame was the list of dwellings in the Labour Force Survey Sample from which households were selected, and the list of individuals of each household. The population is all households from the ten provinces only. The Sample unit is an individual household. A sampling element is an individual. The sampling design involved the collection of data from the household for the household component as well as one individual aged 16 or over in the household for the purpose of the individual component.

Data Collection

Data collection for the Labour Force Survey was carried out each month. Statistics Canada interviewers were hired and trained to carry out the Labour Force Survey and the other household surveys, including the CIUS. Each interviewer contacted 75 dwellings each month. CATI was used to contact households with available telephone numbers and CAPI was used for the rest. CATI accounted for 85% of interviews. Household information was collected from a knowledgeable household member. All interviewers were under supervision of senior interviewers who made sure the interviewers were familiar with the LFS and its supplementary surveys including the CIUS. People who refused to participate, received a letter from the Regional Office stressing the importance of their participation. A second call is then made. If an interviewer's time is inconvenient, an appointment is made. Sampled dwellings were not replaced. For the CIUS, data collection modification included selecting a random individual in the household for the interview. If the respondent was not available, the interviewer called back. Response rate was 81% for 30,817 out of 38039 eligible households for the household component. Individual component yielded a 69% response rate.

Research Questions

We will examine:

1. The relationship between age of respondent and use of ecommerce over the past 12 months.
2. The relationship between recent internet shoppers and concern in using a credit card over the internet.

Variables of interest used in the first relationship examination will be gcagegr6 – Age of respondent and ec_q01 – Over the past 12 months, has the respondent purchased any goods or services over the internet. For the second relationship examination we will look at variables ec_q01 – Over the past 12 months, has the respondent purchased any goods or services over the internet and ps_q02 – Concerned – using credit card over the internet.

Statistical Methods

For the first research question, we will perform Pearson's Chi-Squared test to determine if there is an association between the two variables of interest. The hypothesis test for research question 1 is Ho: Age of respondent and whether the respondent has purchased any goods or services over the internet in the past 12 months are independent. Ha: Age of respondent and whether the respondent has purchased any goods or services over the internet in the past 12 months are dependent. For the second question, we will also conduct a Pearson's Chi-Squared test and find a 95% confidence interval to determine the association between the two variables of interest. The hypothesis test is Ho: Recent internet shoppers and concern of using a credit card over the internet are independent. Ha: Recent internet shoppers and concern of using a credit card over the internet are dependent.

Reduction in Irrelevant Data

```
> table(gcagegr6)
gcagegr6
   1    2    3    4    5    6
1814 3198 3520 4007 4420 5656
> table(ec_q01)
ec_q01
   1    2    6
9384 8226 5005
```

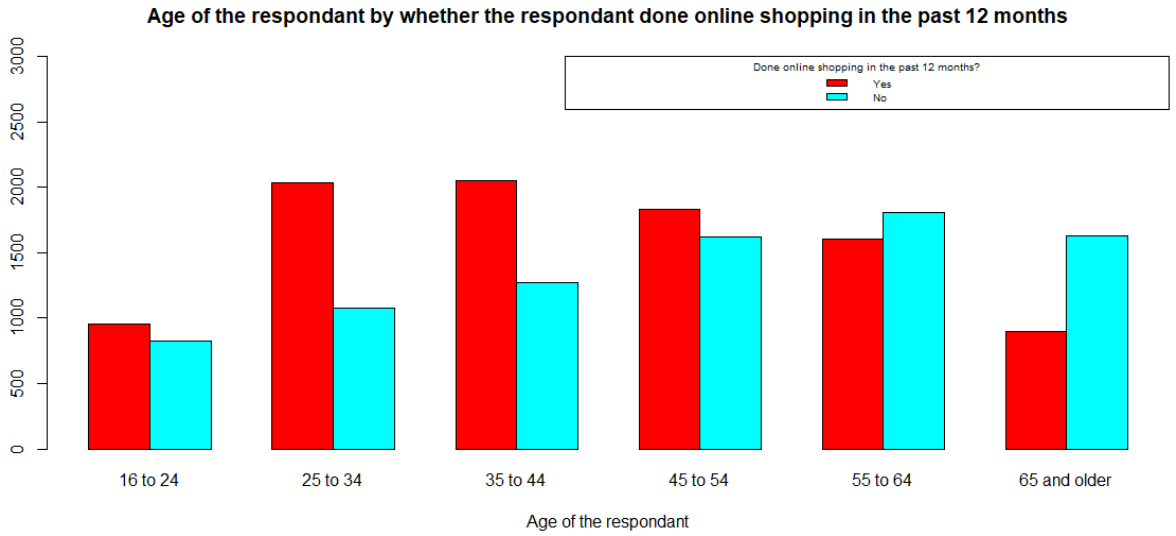
The values of ec_q01 which are 6 indicate a valid skip. These are not relevant in our comparison as we seek to observe whether there is an association between age and online shopping. The values of ps_q02 which are 4 indicate the respondent has no credit card. 6 indicates a valid skip. 7 indicates a don't know response. 8 indicates a refusal response. 9 indicates a not stated response. These are also not relevant as we want to determine whether there is an association between recent internet shoppers and concern with using a credit card over the internet. Removing this irrelevant data gives us:

```
> table(gcagegr6)
gcagegr6
   1    2    3    4    5    6
1783 3112 3324 3459 3411 2521
> table(ec_q01)
ec_q01
   1    2
9384 8226
```

For the first research question. We get a second table for the second research question:

```
> table(ec_q01)
ec_q01
   1    2
9168 7449
> table(ps_q02)
ps_q02
   1    2    3
3731 7424 5462
```

Statistical Analysis for Age of Respondent and Online Shopping Relationship




```
> #Column proportions
> Col.Prop<-prop.table(Table,2)
> Col.Prop
                Age.Of.Respondant
Online.Shopping  16 to 24  25 to 34  35 to 44  45 to 54  55 to 64 65 and olde
r
            Yes 0.5378575 0.6545630 0.6167268 0.5307892 0.4711228    0.355017
9
            No  0.4621425 0.3454370 0.3832732 0.4692108 0.5288772    0.644982
1
```

The bar chart shows a clear relationship between age and recent online shopping. As the age decreases after the "25 to 34" category, so does the use of online shopping. There is a higher proportion of people who recently purchased goods or services online as opposed to those who don't in younger ages while the inverse is apparent in older ages. This is clearly shown in the column proportions table where the percentage of online shoppers rise in younger ages and decreases as ages get older.

Hypothesis Test

We will perform a Chi-squared test of independence to determine whether there is a relationship between the age of the respondent and whether they are a recent online shopper or not.

Ho: Age of respondent and whether the respondent has purchased any goods or services over the internet in the past 12 months are independent.
Ha: Age of respondent and whether the respondent has purchased any goods or services over the internet in the past 12 months are dependent

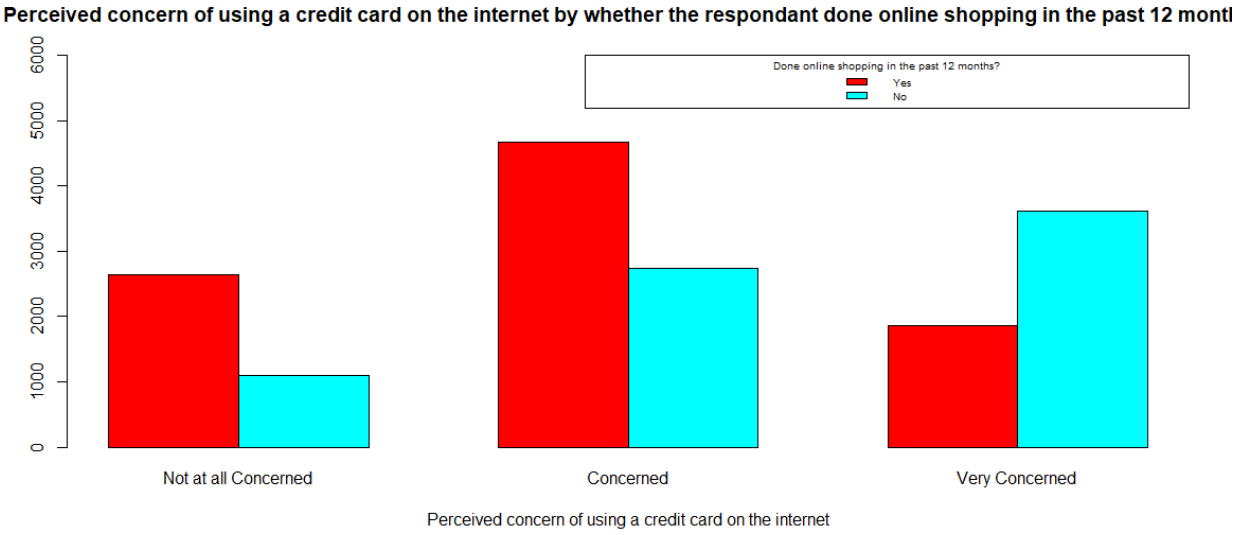
```
> # Perform Chi-Squared test of Independence
> chisq.test(Table)

        Pearson's Chi-squared test

data:  Table
X-squared = 651.89, df = 5, p-value < 2.2e-16
```

Performing this test, we see that we get a p-value of 2.2e-16. Since the p-value is less than 0.05, we can say that we reject the null hypothesis and there is an association between the age of the respondent and whether they have recently done online shopping or not. Based on the bar charts, we can say that as the age of the respondent increases, it is more likely that they have not recently done online shopping.

Statistical Analysis of Age of Respondent and Concern of Credit Card Use Online




```
> # row proportions
> Row.prop<-prop.table(Table,1)
> Row.prop
                CC.Concerned
Online.Shopping Not at all Concerned Concerned Very Concerned
            Yes            0.2874127 0.5103621      0.2022251
            No             0.1471338 0.3685058      0.4843603
```

The bar chart indicates a clear difference in opinion between those who have recently shopped online and those who didn't in terms of concern in the usage of a credit card online. Overall, people who have made purchases online are concerned or not at all concerned with a small amount of people being very concerned while making purchases online using a credit card. Those who have not shopped online recently seem to have a stronger feeling of concern when using a credit card online. Most people who recently shopped online have a feeling of concern when using their credit card online while those who haven't are very concerned. This is seen clearly in the marginal proportion table where the percentage of people who online shop and are concerned in online credit card usage is 51% while those who don't shop online and are very concerned is 48%. In focus of our hypothesis, those who are concerned and have shopped online are concerned at higher numbers than those who have not shopped online.

Hypothesis

We will conduct perform Pearson's Chi-Squared test and create a 95% confidence interval using a 2 sample test for equality of proportions without continuity correction to test our hypothesis:

Ho: Recent internet shoppers and concern of using a credit card over the internet are independent.

Ha: Recent internet shoppers and concern of using a credit card over the internet are dependent.

```
> prop.test(c(2635,1096),c(9168,7449),correct = FALSE)


        Pearson's Chi-squared test

data:  Table
X-squared = 1540.6, df = 2, p-value < 2.2e-16

        2-sample test for equality of proportions without continuity correctio
n

data:  c(2635, 1096) out of c(9168, 7449)
X-squared = 464.48, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 0.1280099 0.1525479
sample estimates:
   prop 1    prop 2
0.2874127 0.1471338
```

Completing the Chi-squared test, we see that we reject Ho since the p value is less than 0.05. Thus, we can say Recent internet shoppers and concern of using a credit card over the internet are dependent. Completing the 2-sample test, we can be 95% confident that the population percentage of those who are recent shoppers on the internet and concerned about using their credit card is 12% to 15% higher than those who are not recent shoppers and concerned about using their credit card online.

Conclusion

We have analyzed the potential relationship between the age of the respondent and online shopping. We have also analyzed the potential relationship between the age of the respondent and concern of credit card use online. For the former, we conducted a Chi squared test on the age of the respondent against whether they have made purchases online or not. Receiving a small p value from the test gives us the conclusion that there is in fact an association between the age of the respondent and whether they have recently done online shopping or not. For the latter, we also conducted a chi squared test which concluded that there was a relationship between the perceived concern of use of a credit card and whether the respondent has done any online shopping in the past 12 months or not.