

Using Google Consumer Surveys

Lars Vilhuber

Contents

Abstract	1
Setup	1
Cost	1
Survey instrument	2
Results	3
Some preliminary conclusions	4
Appendix	5

Abstract

This document reports a simple experiment using Google Consumer Surveys.

Setup

Running surveys over the internet is very popular. Google Consumer Surveys (www.google.com/insights/consumersurveys) is one of those firms that have jumped into the fray, with prices “starting at 10¢ per complete”. We wondered how useful such rapid surveys are for measuring economic concepts. We created a survey, modeled on the [Current Population Survey](#) (CPS), put it into the field for a limited budget, and assessed the results.

```
# define basic parameters
price_minimum_screen <- 3
price_multiq <- 1
price_quoted <- 3
price_minimum <- 0.1
budget <- 500
coupon <- 75
result_file <- c("Employment and search.xls")
survey_start <- as.difftime("11:59",format="%H:%M")
survey_end <- as.difftime("20:24",format="%H:%M")
```

Cost

Google touts that prices are start at \$0.1. That is in fact the price for a single-question, non-targeted survey. For up to 10 questions, the price is \$1 per completed response. In our case, we applied the CPS model and had a screening question, which leads to a minimum price of \$3 per complete response, depending in particular on the incidence rate (a minimum of 5% of respondents must make it through the screener). In our case, the quoted price was \$3 per response, and our budget was \$500 (although we were offered a one-time coupon valued at \$75. We thus requested 191 cases.

Survey instrument

We based our survey instrument on the primary labor force question on the CPS, somewhat modified:

Please complete the following survey to access this premium content.

Question 1 of 2 or fewer:

Last week, did you do ANY work for pay?

☐ No, I am retired


☐ No, I am disabled


☐ No, I was unable to work

☐ Yes

☐ No

OR

 Show me a different question

 Skip survey

Google

[INFO](#) [PRIVACY](#)

We followed up for all non-employed (all those not responding with “Yes”) with a question on search intensity that was not based on the CPS:

Question 2 of 2:

If you are not working, how many jobs did you apply for (by email, website, letter, in person, etc.) last week?

☐ 0 (I am not looking for work)
 ☐ 1-2
 ☐ 3-5
 ☐ 6-10
 ☐ more than 10

OR

Google

[INFO](#)
[PRIVACY](#)

Results

Google first ran a test (which resulted in being 4 responses) to assess how much to charge us. Once accepted, Google posted the survey, targetting the “general population in the United States on the Google Consumer Survey publisher network”. We received 461 cases, of which 192 were “complete”. The entire survey was completed in 8.42 hours. The raw distribution of responses was as follows:

Question.text	Response.c
Last week, did you do ANY work for pay?	
If you are not working, how many jobs did you apply for (by email, website, letter, in person, etc.) last week?	

From this, we can compute a few statistics of interest, which we measured between 2016-01-08 11:10:46 and 2016-01-08 20:20:05.

For comparison purposes, we obtained the BLS unemployment rate and employment-population ratio from

FRED:

```
# define the series of interest
fred_unemployment <- c("UNRATENSA")
fred_unemployment20 <- c("LNU04000024")
fred_emppop_all <- c("LNU02300000")
fred_emppop_men <- c("LNU02300001")
fred_emppop_women <- c("LNU02300002")
fred_olf_count <- c("LNU05000000")
fred_pop_count <- c("CNP160V")

#bls.urate <- get(getSymbols(fred_unemployment,src='FRED'))
bls.urate20 <- get(getSymbols(fred_unemployment20,src='FRED'))
bls.emppop_all <- get(getSymbols(fred_emppop_all,src='FRED'))
#bls.emppop_men <- get(getSymbols(fred_emppop_men,src='FRED'))
#bls.emppop_women <- get(getSymbols(fred_emppop_women,src='FRED'))
#
# Compute the OLF rate
cnt.olf <- get(getSymbols(fred_olf_count,src='FRED'))
cnt.pop <- get(getSymbols(fred_pop_count,src='FRED'))
bls.olf_rate <- cnt.olf*100/cnt.pop
#
# attach the closest value - this might change over time, but that's fine.
# found at http://stackoverflow.com/questions/8186960/finding-the-most-recent-observation-earlier-than-
# For emppop ratio, use all - we will break it out by gender later
results["Employment-Population Rate",2] <-
  round(bls.emppop_all[findInterval(google.dates[1],index(bls.emppop_all)),]
,2)
results["Unemployment rate",2] <-
  round(bls.urate20[findInterval(google.dates[1],index(bls.urate20)),],2)
results["OLF rate",2] <-
  round(bls.olf_rate[findInterval(google.dates[1],index(bls.olf_rate)),],2)
```

The final (unweighted) result is then as follows (note: the BLS numbers won't add up quite to 100, since they stem from different series):

	Google	BLS
Employment-Population Rate	58.35	59.40
Unemployment rate	16.36	4.50
OLF rate	32.10	37.59

Some preliminary conclusions

We probably could have obtained a more informative and cheaper response if not using the labor force question as a screener. Rather, if we had requested search intensity of everybody, regardless of labor force status, our budget would have allowed for 575 responses, *and* given us information on on-the-job search.

Appendix

```
# libraries needed
library(quantmod) # read FRED
library(gdata) # read Excel
options("getSymbols.warning4.0"=FALSE)
Sys.info()
```

```
##                               sysname
##                               "Linux"
##                               release
##                               "3.16.7-29-desktop"
##                               version
## "#1 SMP PREEMPT Fri Oct 23 00:46:04 UTC 2015 (6be6a97)"
##                               nodename
##                               "zotique2"
##                               machine
##                               "x86_64"
##                               login
##                               "vilhuber"
##                               user
##                               "vilhuber"
## effective_user
##                               "vilhuber"
```