Second Amendment to Pre-Analysis Plan:

Analysis of Survey Data

Positive vs. Negative Incentives for Compliance: Evaluating a Randomized Tax Holiday

T. Dunning, F. Monestier, R. Piñeiro, F. Rosenblatt, and G. Tuñón

December 13, 2014

Contents

Description of the document **3**

Challenges to survey implementation **3**

Timeline of survey data collection 5

Revisions to analysis plan **6**

Description of survey experiment 7

Hypotheses: survey experiment 8

Adjustments for multiple comparisons 10

Testing for order effects 11

Revision to Table 7.3 in our original pre-analysis plan 12

Code for mock analysis of survey data **12**

Loading data 12

Correcting outliers and coded vars 13

Shuffling treatment labels 13

t-test function 14

Analysis 15

Social benefits vs individual benefits. 15

Discretionary vs lottery allocation of benefits 17

Fines and charges vs. benefits of tax holidays 22

Description of the document

This document is a second amendment to our original pre-analysis plan, dated July 23, 2014, which is registered at the Experiments in Governance and Politics registration page as well as the study registry of the American Economic Association.[1] We registered our first amendment, dated October 19, 2014, in advance of analyzing administrative (tax payment) data. This second amendment concerns analysis of household survey data, which we have received from our survey firm since filing our first amendment. We register this document in advance of conducting experimental tests using the survey data (though after analyzing descriptive quantities such as unconditional means and standard deviations). We encountered substantial logistical difficulties in implementing aspects of our household survey. Fortunately, these difficulties do not affect our analysis of survey experiments, though they do affect inferences about the effects of our field and natural experiments using the household survey (rather than administrative/tax payment) data.

This document describes our amended plan for analyzing the survey data in light of these considerations. First, we describe the challenges we experienced with respect to survey implementation. Next, we discuss modifications to the analysis plan in light of these challenges. Finally, we present code for a mock analysis of the survey data.

Challenges to survey implementation

We hired a reputable Uruguayan survey firm, CIFRA, to implement our household surveys. We chose CIFRA from among four competing bids due to the reputation of the firm for producing quality work, as well as the lesser cost and shorter promised timeline for the fieldwork relative to other firms.

Our agreement with CIFRA called for face-to-face interviews of 6,000 taxpayers, selected from the total study group of 33,429 households randomly selected from the population of payers of the property tax (Contribución Inmobiliaria) in Montevideo. Fieldwork began July 16, 2014. CIFRA promised to implement our survey in a period of roughly four-five weeks, so that the fieldwork would be completed by the end of August or beginning of September. The timing of the surveys was critical, as we planned to assess the effects on attitudes and behaviors of citizens who received our randomized informational treatments (that is, flyers with different messages about tax payments), which were mailed in phases beginning June 27, 2014 and ending July 16, 2014.

The taxpayer accounts we obtained from the municipality (i.e., our study group, which is in turn a random sample of taxpayers in Montevideo) contained addresses associated to each account. The survey firm intended to locate physical taxpayers using these addresses. There were several considerations regarding the accounts obtained

[1] See registered design 84 [initial date: 20140723] at the EGAP registration page.

from the municipality, which we describe on pages 21-23 of our original pre-analysis plan. First, they contained both physical persons as well as firms ("legal persons" or *personas jurídicas*), since the municipality has no way of identifying ex-ante who is a physical person.[2] Second, the mailing address associated with an account may or may not correspond to the physical address of the property on which taxes are paid.[3] For this reason, it was important for us to work with CIFRA to screen our data and identify households consisting of physical taxpayers.

To accomplish our objective, and in light of several logistical difficulties we encountered, we provided data to CIFRA in two phases:

1. First, as described in our original pre-analysis plan, we randomly sampled 8,000 taxpayer accounts from the study group for our informational field experiment and provided these records to CIFRA, with instructions to seek 6,000 completed surveys from among these. Our procedure required CIFRA to visit each household a minimum of three times, one of which visits had to take place during the weekend. In case of direct refusal or inability of the firm to locate a taxpayer, the firm was to replace that case with another from the original list we provided.[4]

   As of July 25, nine days after initiating fieldwork, the firm had only visited 300 addresses (and only completed 47 interviews). The management of the firm informed us that they had great difficulty locating households, given the primitive mapping technologies they used, and they could not readily organize the fieldwork so as to minimize their costs and expedite the fieldwork. This slow rhythm of work obviously appeared to compromise the firm's ability to complete the surveys in the agreed period.

2. We therefore reorganized the data by geocoding all of the addresses using QGIS together with further municipal data on the location of properties. To do this, we obtained from the municipality the number associated with the physical address of each property on which tax is paid for each account (the *padrón*) and matched this to the official map that assigns the number/*padrón* to a single lot in the city of Montevideo. We then eliminated (for purposes of the household survey) those accounts for which we could not find a valid geolocated address; this reduced the available accounts to 20,866 in the field experiment (from 28,600) and to 4,947 (from 5,129) in the natural experiment (considering only properties subject to property tax or Contribución Inmobiliaria).

[2] Indeed, this is why lottery winners must present themselves to the municipality to establish that they are physical persons and thus eligible for the tax holiday.

[3] In many cases, this did not matter because our informational treatments were mailed to the address associated with the account, and we tracked tax compliance (using administrative data) and attitudes and behaviors (using survey data) using the same address and account data. Thus, we could readily measure the effect of our flyers on a person living at X address who pays taxes on property Y, whether or not X=Y. However, in some cases a renter living in property X might receive the mailer, but the payment of taxes could be made by an absentee landlord (in other cases, the landlord might receive the bill for property X at a different address, and this would be captured in the account data). In other cases, an administrative or real estate agent might receive and pay the bills.

[4] With assistance from us, the firm screened this list of 8,000 to eliminate duplicate entries (e.g. houses that receive more than one bill because they own more than one property) and screen out invalid addresses, leaving a list of 7,107 accounts from which they should select cases for interviews.

Next, we grouped these valid addresses into 931 census tracts. Finally, we randomly sampled 697 census tracts (weighted by the number of households in our study group located in each tract) and sent this first batch of tracts to CIFRA; this batch of census tracts included 25,998 individual addresses. Our hope was that this would dramatically increase the speed with which CIFRA could locate taxpayers in our sample.

Timeline of survey data collection

Unfortunately, our efforts did not have the intended effect. Table 1 shows the progress of the fieldwork conducted by CIFRA. Over the course of eighteen weeks, the firm averaged around 132 successful interviews per week. As of December 11, 2014, CIFRA had completed 2,349 interviews.

Table 0.1: Cumulative progress of survey fieldwork by CIFRA

REPORT DATE	**INTERVIEWS COMPLETED**	HOUSEHOLDS VISITED	Reason for failure			Number of Enumerators
			Refusal	Not physical	Other*	
July 25, 2014	47	300	94	84	75	29
August 5, 2014	240	1,473	449	221	563	53
August 13, 2014	441	2,734	845	350	1,098	49
August 19, 2014	609	3,932	1,088	694	1,541	48
August 26, 2014	821	4,924	1,330	754	2,019	45
September 2, 2014	1,008	6,978	1,876	1,010	3,084	40
September 9, 2014	1,124	8,070	2,283	1,112	3,551	38
September 16, 2014	1,238	9,492	2,627	1,414	4,213	36
September 23, 2014	1,405	11,005	2,998	1,647	4,959	40
September 30, 2014	1,536	12,325	3,426	1,716	5,647	43
October 7, 2014	1,664	13,387	3,784	1,746	6,193	30
October 14, 2014	1,723	14,614	4,024	1,843	7,024	25
October 21, 2014	1,812	15,531	4,326	1,957	7,436	36
October 28, 2014	1,912	16,168	4,429	2,047	7,748	32
November 4, 2014	1,980	16,939	4,719	2,129	8,111	35
November 11, 2014	2,109	17,811	4,938	2,292	8,472	30
November 18, 2014	2,194	18,446	5,120	2,343	8,789	25
November 25, 2014	2,283	18,886	5,202	2,375	9,026	28

* "Others" includes the following reasons for failure: address does not exist; residence is empty; nobody home (after three visits); deceased or repeated address in data set.

This slow rhythm of data collection can potentially compromise inferences from the survey data in two key ways.

1. First—especially for interviews that happened after the intended end of data collection at the end of August—it is perhaps implausible to think that flyers received several weeks or months prior to the interview would substantially impact citizens' responses to the eight questions we registered as main outcomes for our analysis of survey data, namely, (1) trust in municipal government; (2) trust in civil servants; (3) evaluation of the mayor; (4) fairness of municipal taxes in general; (5) fairness of the property tax; (6) attitudes towards tax amnesties;

(7) agreement that sometimes taxes are not worth paying; and (8) party vote intention (or change in vote intention from previous election).[5] Moreover, our survey instrument frequently included temporal references such as “in the last month, have you received a flyer from the municipal government?" which we constructed on the understanding that the survey would be completed in four-five weeks after the mailing of flyers; these references do not make sense for respondents interviewed in September and thereafter.

2. Second, and even more consequentially, power analyses suggested a requisite overall sample size of 6,000 to detect plausible effects of our informational treatments with 80% power, when stratifying between good taxpayers (N=3,000) and bad taxpayers (N=3,000) and estimating effects of individual treatments with respect to the control group. The sample size as of December 11, 2014 is 2,453, that is, less than half of this targeted sample size.

For both reasons, our power to detect effects of our informational treatments using survey data is substantially lower than anticipated by our pre-analysis plan. We do not think it is plausible that we will detect effects of our flyers on the attitudinal outcomes registered in our original plan.

Fortunately, we included in our survey instrument experimental primes that substantially replicate the informational treatments in the field experiment; and we asked several outcome questions that allow us to probe the effects of these primes on political attitudes. Thus, the survey can serve its intended function of measuring the effects of information about punishment for non-payment, or benefits for tax payment (e.g. the tax holiday lottery). We failed to describe the survey experiments in sufficient detail in our original pre-analysis plan and thus do so here.

Revisions to analysis plan

In light of the complications that arose during survey data collection, and our reliance on the survey experiment to draw inferences about the effects of informational treatments, we make the following amendments to our plan for analysis of survey data:

1. We consider the survey measures registered in our field and natural experiments (see page 30 of our original pre-analysis plan) to be secondary outcomes. Our primary outcomes for the field and natural experiments are the administrative measures of tax compliance, missed payments, and total debt, as well as (for the field experiment) our measure of Web bill access (see page 29 of our original plan). We will adjust statistical tests for multiple comparisons with respect to those primary outcomes measured through administrative data.

[5] See page 30 of our original pre-analysis plan.

2. We will focus the analysis of survey data on the survey experiments (see previous section and page 30-31 of our original plan). Note that these questions are not subject to the same issues regarding delays in the surveys, because randomization of treatments is achieved within the survey itself. Analysis of these questions does not give us purchase on the effects of our informational flyers but does serve some of the same objectives, in terms of giving us insight into how priming sanctions/punishments vs. rewards may influence attitudes.

3. Finally, we will use the survey data to gain insight into observational quantities such as the proportion of taxpayers in Montevideo who know about the existence of the tax holiday lottery, or the existence of tax amnesties as well as fines and punishment for non-payment. These analysis should be taken with some grain of salt as appropriate, given the substantial non-response documented in Table 1 (i.e., we may only project to the population of taxpayers in Montevideo under some strong assumptions). But these data will still be useful for assessing reasons why the natural or field experimental treatments did not have a larger effect on tax payment behavior (as measured through the field experiment).

Description of survey experiment

Our survey instrument included the following informational treatments; the order was randomized and, as we describe below, individuals were randomly assigned to receive certain of these prompts and not others.

- INDIVIDUAL PUNISHMENT: "The municipal government of Montevideo applies fines and charges to those who do not pay their taxes on time. These punishments can be very costly for whoever does not pay their taxes."[6]

- SOCIAL PUNISHMENT: "The municipal government of Montevideo applies fines and charges to those who do not pay their taxes on time. Fines and charges are a punishment for those who do not contribute to the construction of a city that is more just and better for all."[7]

- INDIVIDUAL BENEFIT: "The municipal government of Montevideo raffles, in conjunction with the National Lottery, the exoneration of municipal taxes for one year. This policy individually rewards those who are up to date on their tax payments."[8]

- SOCIAL BENEFIT: "The municipal government of Montevideo raffles, in conjunction with the National Lottery, the exoneration of municipal taxes for one year among people who are up to date on their tax payments. The

[6] In Spanish, "La Intendencia de Montevideo aplica multas y recargos a quienes no pagan sus impuestos en fecha. Estas sanciones pueden ser muy costosas para quien no paga sus impuestos."

[7] La Intendencia de Montevideo aplica multas y recargos a quienes no pagan sus impuestos en fecha. Las multas y recargos son una sancin para quienes no contribuyen a la construccin de una ciudad ms justa y mejor para todos y todas.

[8] La Intendencia de Montevideo sortea junto a la Lotera Nacional la exoneracin de los tributos municipales por un ao. Esta poltica premia individualmente a quienes estn al da con sus impuestos.

municipal government conducts this lottery to recognize good taxpayers for their contribution to the construction of a city that is more just and better for all.[9]

- "DISCRETIONAL" BENEFIT: "At times the municipal government of Montevideo chooses people who are up to date on their municipal tax payments, to grant them the exoneration of municipal taxes for one year."[10]

Hypotheses: survey experiment

We will test several hypotheses about the effects of these informational treatments on several of the attitudinal outcome measures registered in our original pre-analysis plan (see p. 30-31). Respondents were asked for their degree of agreement on a 0-10 scale with the following statements:

1. "People only pay their taxes on time when there are substantial fines and charges" (survey question M.1.1);[11]
2. "In Montevideo, punishments don't apply to the privileged" (question M.1.4)[12]; and
3. "Fines and charges for bad taxpayers are pointless" (question M.1.5).[13]
4. "Policies that reward good taxpayers are a waste of money" (question S.1.1)[14];
5. "In Montevideo, benefits for good taxpayers go to the same people as always (question S.1.4)."[15]
6. "In general, the municipal government does a good job" (questions M.1.3 and S.1.2)[16];
7. "In Montevideo, it is worth it to be up to date on ones taxes" (question M.1.2 and S.1.3)"[17]

   as well as their response to a question with four ordered response categories:

8. "How would you classify the taxes that the municipal government charges, in general: very just, fairly just, a little just, or not just at all?" (questions M.1.6 and S.1.5)[18]

We reiterate and slightly modify hypotheses outlined in our original pre-analysis plan (see e.g. Table 7.3). In particular, we are interested in the testing the alternative of a treatment effect against the following null hypotheses:

[9] La Intendencia de Montevideo sortea junto a la Lotera Nacional la exoneracin de los impuestos por un ao entre las personas que estn al da. La Intendencia realiza este sorteo para reconocer a los buenos pagadores por su contribucin a la construccin de una ciudad ms justa y mejor para todos/as.
[10] A veces la Intendencia de Montevideo elige personas que estn al da con el pago de impuestos municipales, para otorgarles un ao de exoneracin de pago de estos impuestos.
[11] In Spanish, "La gente paga sus impuestos en fecha solo cuando las multas y los recargos son importantes."
[12] In Spanish, "En Montevideo, a los privilegiados, no se les aplican los castigos."
[13] In Spanish, "Las multas y recargos a malos pagadores no sirven para nada."
[14] In Spanish, "Las políticas que premian a los buenos pagadores son un derroche de dinero."
[15] In Spanish, "En Montevideo, los beneficios para buenos pagadores se los llevan los mismos de siempre."
[16] In Spanish, "En general, la Intendencia hace un buen trabajo."
[17] In Spanish, "En Montevideo, vale la pena estar al día con los impuestos."
[18] In Spanish, "Cómo clasificaría los impuestos que cobra la Intendencia de Montevideo en general: muy justos, bastante justos, poco justos o nada justos?"

1. PUNISHMENTS VERSUS BENEFITS: There is no difference in the response of taxpayers to the benefit treatments vs. the punishment treatments, for the following outcomes[19]:

   - "In general, the municipal government does a good job" (question S.1.2 and question M.1.3)[20];

     Here, we expect benefits to have a larger positive effect than punishments (where the effect may be negative), and thus we expect benefits - punishments to be positive; we will conduct a one-tailed test consistent with this alternative hypothesis.

   - "In Montevideo, it is worth it to be up to date on ones taxes" (question M.1.2 and question S.1.3);

     Here, we have no strong directional hypothesis: either punishments or benefits may have the larger effect on this variable. We will conduct a two-tailed test.

   - "How would you classify the taxes that the municipal government charges, in general: very just, fairly just, a little just, or not just at all?" (questions M.1.6 and S.1.5); and

     Here, we again expect benefits to have the larger positive effect (i.e. greater perceptions that taxes are just), leading to a one-tailed test.

2. DISCRETIONARY VS. NON-DISCRETIONARY/RANDOM BENEFITS: There is no difference in the response of taxpayers to the discretionary benefit vs. individual/social benefits conditions in which we mention the lottery, for the following outcomes[21]:

   - "In Montevideo, it is worth it to be up to date on ones taxes" (question S.1.3);

     Here, we suspect the non-discretionary benefits (knowledge of lottery) will have the larger positive effect, as we only state that rewards are conditional on being up to date (being a "good taxpayer") in the non-discretionary/random benefits conditions. We will conduct a one-tailed test.

   - "Policies that reward good taxpayers are a waste of money" (question S.1.1);

     We conduct a two-tailed test, because taxpayers might dislike discretion, but they might also dislike lotteries if they perceive them to be ineffective.

   - "In general, the municipal government does a good job" (question S.1.2);

     We conduct a two-tailed test, because taxpayers might like "universal particularism" (Monestier et al. 2014) but they might also like the lottery policy.

   - "In Montevideo, benefits for good taxpayers always go to the same people" (question S.1.4); and

[19] This corresponds to Hypothesis 2A in the original plan (see Table 7.3), and we add one outcome not registered there

[20] Question S.1.2. is the same as M.1.3; however, the former follows the prompt about benefits whereas the latter follows the prompts about punishments. A similar comment applies to other outcomes for the PUNISHMENT VERSUS BENEFITS hypotheses.

[21] This corresponds to Mechanisms IA.1 and IA.2 in the original plan (see Table 7.3).

We conduct a one-tailed test under the alternative hypothesis that the discretionary benefits treatment increases agreement with this statement more than the lottery treatment.

- "How would you classify the taxes that the municipal government charges, in general: very just, fairly just, a little just, or not just at all?" (question S.1.5).

  We conduct a one-tailed test under the alternative hypothesis that the non-discretionary/lottery treatment increases perceptions that the system is just.

3. SOCIAL VERSUS INDIVIDUAL BENEFITS: There is no difference in the response of taxpayers to the social benefit and the individual benefit primes, for the following outcomes[22]:

   - "Policies that reward good taxpayers are a waste of money" (question S.1.1); and

     We conduct a two-tailed test, because who the heck knows? (Certainly not us. That's why we are doing the research!).

   - "In Montevideo, it is worth it to be up to date on ones taxes" (question S.1.3).

     The treatment effects depend on whether people are rational individual maximizers or social beings. We are agnostic. Two-tailed test.

Adjustments for multiple comparisons

Our original pre-analysis plan did not specify corrections for the survey experimental analysis in detail, so we do so here. As per our original plan, we will present both nominal p-values and corrected p-values, using a false discovery rate (FDR) correction to control the Type-1 error rate. For the survey experiments, as for our natural and field experiments, we will control the FDR at level 0.05.

For a given randomization with m (null) hypotheses and m associated p-values, we order the realized nominal p-values from smallest to largest, $p_{(1)}, \leq p_{(2)} \leq \ldots \leq p_{(m)}$. Let

$$k \text{ be the largest } i \text{ for which } p_{(i)} \leq \tfrac{i}{m} 0.05.$$

Then, we reject all $H_{(i)}$ for $i = 1, 2, \ldots, k$, where $H_{(i)}$ is the null hypothesis corresponding to $p_{(i)}$.[23]

For comparison, we will also present strict Bonferroni corrections, i.e., for each hypothesis $H_{(i)}$, we reject the null at the adjusted 0.05 level if $p_{(i)} \leq \frac{0.05}{m}$. This correction will lead to the most conservative inference for each

[22] Though we specified a comparison of social vs. individual benefits for the field experiment in our original plan, we omitted this comparison of survey experimental treatments; we add that here.

[23] For a description of this procedure, see Yoav Benjamini and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological).* 57 (1): 289-300.

individual pairwise comparison. Our rejection rule, however, will require controlling the overall false discovery rate.

How large is m under our study design? This differs for the survey experiment, we have three independent or nearly independent set of tests corresponding to different randomizations. For each set of tests, there are two treatment conditions. The number of outcome variables/test statistics (and associated p-values) in each set is as follows:

- PUNISHMENT VS. BENEFITS: $m = 3$
- DISCRETIONARY BENEFITS VS. NON-DISCRETIONARY BENEFITS/LOTTERY: $m = 5$
- SOCIAL VS. INDIVIDUAL BENEFITS: $m = 2$.

Testing for order effects

Each survey respondent was exposed to two of the informational treatments (though in different orders, as explained next). For purposes of our main analysis, we pool treatments, e.g. for H1 we compare all exposures to the Social Benefit prime to all exposures to the Individual Benefit prime; and for H2 we pool the social and individual benefit treatments and compare them to the pooled social and individual punishment conditions.

To allow us to test for order effects, we varied the placement of the treatments across four versions of the questionnaires, as shown in Table 2. The "Discretional Benefit" treatment only appears in one questionnaire (D) and appears before the other informational prime in that questionnaire; thus, it cannot be subject to order effects. To test whether the other three primes are subject to order effects, we will assess whether responses to the individual benefit, individual punishment, and social punishment conditions are statistically identical when these treatments appear as the first informational treatment vs. as the second informational treatment in the questionnaire. Note that each of these three treatments appears once in the first position on a questionnaire version and once in the second position. Thus, even if there are order effects, they should not bias conclusions about the effects of these treatments compared to each other.

Table 0.2: Survey experimental treatments: distribution across four versions of questionnaire

	Version of Questionnaire			
	A	B	C	D
First treatment	Individual Punishment	Social Punishment	Individual Benefit	Discretional Benefit
Second treatment	Social Benefit	Individual Benefit	Individual Punishment	Social Punishment

Revision to Table 7.3 in our original pre-analysis plan

In light of the modifications discussed in this document, we amend the set of hypotheses, outcomes and tests described in Table 7.3 of our original pre-analysis plan, mainly in order to exclude tests in our field experiment using survey data. Here, strikethroughs indicate tests that we will not longer undertake. In a few cases, we add outcomes for the survey experiment we had forgotten to register in the original plan. We also update this table to reflect changes to the analysis of the administrative/tax payment data registered in our first amendment to the pre-analysis plan (which we filed before analyzing the administrative data).

Code for mock analysis of survey data

In what follows, we use a dataset containing responses from 2,349 survey respondents, provided to us by CIFRA on December 11, 2013. We will replace this with the final dataset containing 2,400 respondents when we receive it from the survey firm.

Here, after cleaning some variables where we detected coding errors (e.g. as indicated by extreme/impossible outliers), we shuffle treatment labels, that is, we assign each respondent at random to a version of the questionnaire (and thus of survey experimental prompts) that does not necessarily reflect the real version/prompts that the respondent received. This allows us to run code for our survey experimental tests without seeing the real results. We created the commented code below in R and generated a TeX file using R markdown.

```
rm(list=ls())
library(foreign)
library(ggplot2)
```

Loading data

```
setwd("~/Dropbox/Uruguay state capacity/C722MDEO 11-12-2014")


cA <- read.csv("C722MDEOA.csv")
cB <- read.csv("C722MDEOB.csv")
cC <- read.csv("C722MDEOC.csv")
cD <- read.csv("C722MDEOD.csv")
```

```
data <- rbind(cA,cB,cC,cD)
```

Correcting outliers and coded vars

```
data$B1[data$B1==7070] <- NA
data$B1[data$B1==87] <- 0
data$B2[data$B2==87] <- 0
data$S4[data$S4==7] <- "Otro"
data$D11[data$D11==7] <- "Otro"
data$D11p2[data$D11p2==5] <- "Otro"
data$E3[data$E3==99] <- "No voto"
data$E4p2[data$E4p2==99] <- "No voto"
data$E5[data$E5==99] <- "No vota en Montevideo"
data$E6p2[data$E6p2==99] <- "No simpatiza con ningun partido"
data$D9p3[data$D9p3==0] <- NA # confirmar con CIFRA
data$G3[data$G3==8] <- NA

# Replacing 88, 89, 99 as NA
for (i in 3:ncol(data)){
  data[,i][data[,i]==88] <- NA
  data[,i][data[,i]==89] <- NA
  data[,i][data[,i]==99] <- NA
}
```

Shuffling treatment labels

```
data$treatment <- sample(data$tipo, length(data$tipo), replace=F)
```

When we are ready to analize the real data we will replace the line above with the following line:

```r
# data$treatment <- data$tipo
```

t-test function

```r
# t test with SEs
# df for two sample t test with unequal variances
t.test.se <- function(y,x, two.tailed=TRUE){

  # Calculating difference in means
  mean1 <- mean(y[x==1], na.rm=T)
  mean0 <- mean(y[x==0], na.rm=T)
  diff <- mean1 - mean0

  # Calculating SE of the difference
  N1 <- length(na.omit(y[x==1]))
  N0 <- length(na.omit(y[x==0]))
  var1 <- var(y[x==1],na.rm=T)
  var0 <- var(y[x==0],na.rm=T)
  varN1 <- var1/N1
  varN0 <- var0/N0
  se.diff <- sqrt(varN1 + varN0)

  # T-statistic
  t <- diff/se.diff

  # Degrees of freedom
  df.num <- ((varN1 + varN0)^2)
  df.den <- (varN1^2)/(N1-1) + (varN0^2)/(N0-1)
  df <- df.num/df.den

  # P-value
  if(two.tailed==TRUE){
```

```
        if (t>=0) { p <- pt(t, df, lower.tail=F) +
                      pt(-t, df, lower.tail=T) }
        if (t<0) {p <- pt(t, df, lower.tail=T) +
                    pt(-t, df, lower.tail=F)}
    }

  if(two.tailed==FALSE){
        if (t>=0) {p <- pt(t, df, lower.tail=F)}
        if (t<0) {p <- pt(t, df, lower.tail=T)}
     }

  # Preparing output
  res <- c(mean1, mean0, diff, se.diff,
           t, (N1+N0), df, p)
  names(res) <- c("Mean 1", "Mean 0", "Difference",
                  "SE Diff","t-stat", "N", "df", "p-value")

  return(c(res))
}
```

Analysis

Social benefits vs individual benefits.

```
# Pooling the individual benefits treatments (questionaire versions 2 and 3)
# and recoding as treatment=1.
data$social_individual <- ifelse((data$treatment==2|data$treatment==3),1,NA)
# Recoding the social benefits treatment as 0.
data$social_individual <- ifelse((data$treatment==1),0,data$social_individual)

# Outcome: "Policies that reward good taxpayers are a waste of money"
# totally disagree (0) - totally agree (10)
```

```
ben1 <- with(data,
     t.test.se(S1p1, social_individual, two.tailed=TRUE))

# Outcome: "It is worth it to be up to date with ones taxes"
# totally disagree (0) - totally agree (10)
ben2 <- with(data,
     t.test.se(S1p3, social_individual, two.tailed=TRUE))

social_individual <- rbind(ben1, ben2)
rownames(social_individual) <- c("Rewards are waste of money",
                                 "Worth it to be up to date")
```

```
social_individual <- social_individual[order(social_individual[,8], decreasing=F),]

# Ordering p-values in decreasing order
ordered.ps <- social_individual[, 8]

# Building reference vector to compare to ordered p-values
FDR_reference <- .05*(1:length(ordered.ps)/length(ordered.ps))

# Comparing p-values to referece vector
FDR <- as.data.frame(cbind(ordered.ps,FDR_reference,
                           ordered.ps<=FDR_reference))

if (sum(FDR[,3])>0){
  fdr <- which(FDR[,1]==max(FDR[,1][FDR[,3]==1]))
  FDR[,4] <- c(rep("reject null", fdr), rep("do not reject", nrow(FDR)-fdr))}
if (sum(FDR[,3])==0){
  FDR[,4] <- rep("do not reject", nrow(FDR))}

Bonferroni_reference <- rep(.05/nrow(FDR), nrow(FDR))
```

```
Bonferroni_reject <- ifelse(ordered.ps<=Bonferroni_reference,
                            "reject null", "do not reject")

social_individual <- cbind(social_individual,
                           FDR[,c(2,4)],
                           Bonferroni_reference,
                           Bonferroni_reject)

names(social_individual)[10] <- "FDR_reject"

social_individual
```

Adjusting p-values

```
##                            Mean 1 Mean 0 Difference SE Diff  t-stat    N
## Rewards are waste of money  3.130  3.002    0.12840  0.1849  0.6946 1660
## Worth it to be up to date   7.412  7.509   -0.09638  0.1537 -0.6269 1682
##                              df p-value FDR_reference    FDR_reject
## Rewards are waste of money 1175  0.4875         0.025 do not reject
## Worth it to be up to date  1199  0.5308         0.050 do not reject
##                            Bonferroni_reference Bonferroni_reject
## Rewards are waste of money                0.025     do not reject
## Worth it to be up to date                 0.025     do not reject
```

Discretionary vs lottery allocation of benefits

```
# Generating discretionary benefits dummy where surveys with the discretionary version
# of the survey are 1.
data$treat_discretion <- ifelse((data$treatment==4),1,0)

# Outcome: "In Montevideo, rewards for good taxpayers go to the same people as always"
# totally disagree (0) - totally agree (10)
```

```
discretion1 <- with(data,
     t.test.se(S1p4, treat_discretion, two.tailed=FALSE))
discretion1
```

```
##     Mean 1     Mean 0 Difference    SE Diff     t-stat          N
##  5.310e+00  5.293e+00  1.665e-02  2.211e-01  7.534e-02  1.542e+03
##          df    p-value
##  7.068e+02  4.700e-01
```

```
# Outcome: "Policies that reward good taxpayers are a waste of money"
# totally disagree (0) - totally agree (10)
discretion2 <- with(data,
     t.test.se(S1p1, treat_discretion, two.tailed=TRUE))
discretion2
```

```
##     Mean 1     Mean 0 Difference    SE Diff     t-stat          N
##  3.108e+00  3.086e+00  2.187e-02  1.789e-01  1.222e-01  2.234e+03
##          df    p-value
##  9.674e+02  9.027e-01
```

```
# Outcome: "It is worth it to be up to date with ones taxes"
# totally disagree (0) - totally agree (10)
discretion3 <- with(data,
     t.test.se(S1p3, treat_discretion, two.tailed=FALSE))
 discretion3
```

```
##     Mean 1     Mean 0 Difference    SE Diff     t-stat          N
##  7.736e+00  7.445e+00  2.910e-01  1.341e-01  2.169e+00  2.266e+03
##          df    p-value
##  1.123e+03  1.513e-02
```

```
# Outcome: "In general, the municipal government does a good job"
# totally disagree (0) - totally agree (10)
discretion4 <- with(data,
```

```
     t.test.se(S1p2, treat_discretion, two.tailed=TRUE))
discretion4
```

```
##     Mean 1     Mean 0 Difference    SE Diff     t-stat          N
##     4.7785     4.5649     0.2136     0.1369     1.5602  2313.0000
##         df    p-value
##  1102.6385     0.1190
```

```
# Outcome: "How would you classify the taxes that the municipal government charges in general"
# very just (1) - not just at all (4)
discretion5 <-with(data,
     t.test.se(S1p5, treat_discretion, two.tailed=FALSE))
discretion5
```

```
##     Mean 1     Mean 0 Difference    SE Diff     t-stat          N
##    2.66497    2.74222   -0.07725    0.03420   -2.25910 2291.00000
##         df    p-value
## 1073.84987    0.01204
```

```
discretion <- rbind(discretion1, discretion2, discretion3,
                    discretion4, discretion5)
rownames(discretion) <- c("Rewards go to the same people as always",
                          "Rewards are waste of money",
                          "Worth it to be up to date",
                          "Mun.gov. does a good job",
                          "Mun. taxes are just")
```

```
discretion <- discretion[order(discretion[,8], decreasing=F),]

# Ordering p-values in decreasing order
ordered.ps <- discretion[, 8]
```

```r
# Building reference vector to compare to ordered p-values
FDR_reference <- .05*(1:length(ordered.ps)/length(ordered.ps))

# Comparing p-values to referece vector
FDR <- as.data.frame(cbind(ordered.ps,FDR_reference,
                           ordered.ps<=FDR_reference))

if (sum(FDR[,3])>0){
  fdr <- which(FDR[,1]==max(FDR[,1][FDR[,3]==1]))
  FDR[,4] <- c(rep("reject null", fdr), rep("do not reject", nrow(FDR)-fdr))}
if (sum(FDR[,3])==0){
  FDR[,4] <- rep("do not reject", nrow(FDR))}

Bonferroni_reference <- rep(.05/nrow(FDR), nrow(FDR))

Bonferroni_reject <- ifelse(ordered.ps<=Bonferroni_reference,
                            "reject null", "do not reject")

discretion <- cbind(discretion,
                           FDR[,c(2,4)],
                           Bonferroni_reference,
                           Bonferroni_reject)

names(discretion)[10] <- "FDR_reject"

discretion
```

Adjusting p-values

```
##                                    Mean 1 Mean 0 Difference SE Diff
## Mun. taxes are just                 2.665  2.742   -0.07725  0.0342
## Worth it to be up to date           7.736  7.445    0.29100  0.1341
## Mun.gov. does a good job            4.779  4.565    0.21358  0.1369
```

```
## Rewards go to the same people as always  5.310  5.293    0.01665  0.2211
## Rewards are waste of money               3.108  3.086    0.02187  0.1789
##                                          t-stat    N     df p-value
## Mun. taxes are just                     -2.25910 2291 1073.8 0.01204
## Worth it to be up to date                2.16933 2266 1123.2 0.01513
## Mun.gov. does a good job                 1.56022 2313 1102.6 0.11899
## Rewards go to the same people as always  0.07534 1542  706.8 0.46998
## Rewards are waste of money               0.12223 2234  967.4 0.90274
##                                         FDR_reference    FDR_reject
## Mun. taxes are just                              0.01   reject null
## Worth it to be up to date                        0.02   reject null
## Mun.gov. does a good job                         0.03 do not reject
## Rewards go to the same people as always          0.04 do not reject
## Rewards are waste of money                       0.05 do not reject
##                                         Bonferroni_reference
## Mun. taxes are just                                     0.01
## Worth it to be up to date                               0.01
## Mun.gov. does a good job                                0.01
## Rewards go to the same people as always                 0.01
## Rewards are waste of money                              0.01
##                                         Bonferroni_reject
## Mun. taxes are just                         do not reject
## Worth it to be up to date                   do not reject
## Mun.gov. does a good job                    do not reject
## Rewards go to the same people as always     do not reject
## Rewards are waste of money                  do not reject
```

Conditional on finding significant p-values ($p<.05$) for either of these tests, we will test for order effects by comparing the mean of the "individual benefits" treatment when it appears first to its mean when it appears second to see if they are statistically different.

Fines and charges vs. benefits of tax holidays

```
# Creating dataframe with all treatments and outcomes.


# Keeping outcomes we want for benefits
ben <- data[(data$treatment!=4),c(59,60,62,111)]
names(ben)
```

```
## [1] "S1p2"      "S1p3"      "S1p5"      "treatment"
```

```
# Benefits pooled (S1p2) (A, B and C)
ben$benefits_punishments <- 1


# versus fines and charges pooled (M1p3)  (A, B , C and D)
fin <- data[,c(50,51,54,111)]
names(fin)
```

```
## [1] "M1p2"      "M1p3"      "M1p6"      "treatment"
```

```
# Pooling punishments
fin$benefits_punishments <- 0


# For three questions the outcomes are the same in the punishments and benefits conditions
# but have different survey question numbers. Here we rename the variables so
# that we can bind the datasets into one.
names(ben)
```

```
## [1] "S1p2"                 "S1p3"                 "S1p5"
## [4] "treatment"            "benefits_punishments"
```

```
names(ben)[1:3] <- c("M1p3", "M1p2", "M1p6")


pooled <- rbind(ben,fin)
```

```r
# Checking that the dimensions of the pooled dataset are correct.
stopifnot(nrow(data)*2-nrow(cD)==nrow(pooled))
```

```r
# Outcome: "In general, the municipal government does a good job"
# totally disagree (0) - totally agree (10)
benefits_punishments1 <- with(pooled,
     t.test.se(M1p3, benefits_punishments, two.tailed=FALSE))
benefits_punishments1
```

```
##     Mean 1     Mean 0 Difference    SE Diff     t-stat          N
##    4.56494    4.76322   -0.19828    0.09662   -2.05223 4006.00000
##         df    p-value
## 3694.54263    0.02011
```

```r
# Outcome: "In Montevideo, it is worth it to be up to date on ones taxes"
# totally disagree (0) - totally agree (10)
benefits_punishments2 <- with(pooled,
     t.test.se(M1p2, benefits_punishments, two.tailed=TRUE))
benefits_punishments2
```

```
##     Mean 1     Mean 0 Difference    SE Diff     t-stat          N
##  7.445e+00  7.853e+00 -4.082e-01  9.419e-02 -4.334e+00  3.955e+03
##         df    p-value
##  3.459e+03  1.508e-05
```

```r
# Outcome: "How would you classify the taxes that the municipal government charges?"
# very just (1) - not just at all (4)
benefits_punishments3 <- with(pooled,
     t.test.se(M1p6, benefits_punishments, two.tailed=FALSE))
benefits_punishments3
```

```
##     Mean 1     Mean 0 Difference    SE Diff     t-stat          N
##  2.742e+00  2.705e+00  3.729e-02  2.377e-02  1.569e+00  3.994e+03
##         df    p-value
```

```
##  3.651e+03  5.841e-02
```

```
benefits_punishments <- rbind(benefits_punishments1,
                              benefits_punishments2,
                              benefits_punishments3)
rownames(benefits_punishments) <- c("Mun. gov. does a good job",
                                    "Worth it to be up to date",
                                    "Mun. taxes are just")
```

```
benefits_punishments <- benefits_punishments[order(benefits_punishments[,8], decreasing=F),]

# Ordering p-values in decreasing order
ordered.ps <- benefits_punishments[, 8]

# Building reference vector to compare to ordered p-values
FDR_reference <- .05*(1:length(ordered.ps)/length(ordered.ps))

# Comparing p-values to referece vector
FDR <- as.data.frame(cbind(ordered.ps,FDR_reference,
                           ordered.ps<=FDR_reference))

if (sum(FDR[,3])>0){
  fdr <- which(FDR[,1]==max(FDR[,1][FDR[,3]==1]))
  FDR[,4] <- c(rep("reject null", fdr), rep("do not reject", nrow(FDR)-fdr))}
if (sum(FDR[,3])==0){
  FDR[,4] <- rep("do not reject", nrow(FDR))}

Bonferroni_reference <- rep(.05/nrow(FDR), nrow(FDR))

Bonferroni_reject <- ifelse(ordered.ps<=Bonferroni_reference,
                            "reject null", "do not reject")
```

```
benefits_punishments <- cbind(benefits_punishments,
                              FDR[,c(2,4)],
                              Bonferroni_reference,
                              Bonferroni_reject)

names(benefits_punishments)[10] <- "FDR_reject"

benefits_punishments
```

Adjusting p-values

```
##                            Mean 1 Mean 0 Difference SE Diff t-stat    N
## Worth it to be up to date  7.445  7.853   -0.40819 0.09419 -4.334 3955
## Mun. gov. does a good job  4.565  4.763   -0.19828 0.09662 -2.052 4006
## Mun. taxes are just        2.742  2.705    0.03729 0.02377  1.569 3994
##                             df   p-value FDR_reference    FDR_reject
## Worth it to be up to date 3459 1.508e-05       0.01667   reject null
## Mun. gov. does a good job 3695 2.011e-02       0.03333   reject null
## Mun. taxes are just       3651 5.841e-02       0.05000 do not reject
##                            Bonferroni_reference Bonferroni_reject
## Worth it to be up to date              0.01667       reject null
## Mun. gov. does a good job              0.01667     do not reject
## Mun. taxes are just                    0.01667     do not reject
```

Conditional on finding significant p-values ($p<.05$) for either of these tests, we will test for order effects by comparing the mean of the "individual benefits", "individual punishment" and "social punishment" treatments when they appear first to their mean when they appear second to see if they are statistically different.