Randomized experiments, the NSW program

1 Background information on NSW

Some groups of people have trouble being inserted in the labor market. Examples are people who were released from prison and former drug-addicts. The professional insertion of these groups may be desirable in society, yet is not an easy task to implement. Programs funded by government to train and insert individuals in the labor market are costly, and their effectiveness is often subject to political debate, particularly during budget tightening periods.

In the mid-1970s in the United States, a program called National Supported Work Demonstration (NSW) was launched to support disadvantaged groups to be inserted in the labor market. This program provides an ideal opportunity for a quantitative impact assessment of its effectiveness, because the program randomized its interventions among a population of eligible individuals. In the program, individuals were offered employment with subsidized wages for a period of 12 to 18 months. The program benefited several categories of disadvantaged groups, such as poor women with children (women who had been receiving Aid to Families with Dependent Children), ex-drug addicts, ex-sexual offenders, and young school dropouts, often with criminal records or history of delinquency. The NSW was designed to test whether this short period of assisted employment enhanced the professional outcomes of the people who took it.

2 Data

In the dataset, you have the following variables:
treat : indicator 1 if the person was in the treatment group, 0 if in the control group
age: age of the person
education: number of years of education
black: indicator, 1 if the person Afro-American
hispanic: indicator, 1 if the person is Hispanic
married: indicator, 1 if the person is married
nodegree: indicator, 1 if the person is a high school dropout
re75: income in 75, before the program
re78: income in 78, after the program
ethnic: 1 if the person is white, 2 if Afro-American and 3 if Hispanic

3 Questions

1) Short questions

1. What economic question is this program trying to answer? The question is to measure the effect of a supported job training program on the employment opportunities and earnings of groups that are typically hard to employ.

2. What is the observation level? The observation level in the training program dataset is the individual (i.e. each row is a person).

3. What is the randomization level? The randomisation level is also the individual (i.e. nationally, individuals were randomly selected to take the job training program).

4. What is the intervention variable, and what is the outcome variable? The intervention variable is whether a person participated or not in the program (in the dataset this is variable *treat*), and the outcome variable on which we measure the impact of the intervention is the post-program income of the person (encoded in the variable *re78*).

2) Why would you randomize at the individual level or at a more aggregated level (such as stratified randomization)?
Think about the design of the intervention : other levels of randomization are possible, on zipcode for instance. Why would you do that? Randomization at the individual level (i.e. the most granular level) is the ideal setting. Sometimes the randomization is done at a coarser level when we cannot control individual access to a treatment (town facility for instance) or when we suspect contagion from the treatment to the control group (when there are peer effects or equilibrium effects). For example Progresa was randomized at the village level to avoid such general equilibrium effects or peer effects.

An alternative to randomization at the individual level is the stratified randomization method. The stratified randomization method addresses the need to control and balance the influence of covariates. Specific covariates must be identified beforehand by the researcher, based on a theoretical understanding of the potential influence that each covariate has on the dependent variable. Stratified randomization is achieved by generating a separate block for each combination of covariates. After all subjects have been identified and assigned into blocks, simple randomization is performed within each block to assign subjects to one of the groups.

Its advantages include minimizing sample selection bias and ensuring certain segments of the population are not overrepresented or underrepresented. In our case, if the level of education affects the income, we want to make sure that our treatment effect is not driven by an imbalance of the treatment groups.

3) What is the causal effect you are looking for? What causal effect can you hope to identify with randomization?
The causal or treatment effect is the difference between the two potential outcomes of an individual: the outcome when treated (call it $y_i(1)$ and the outcome when not treated $y_i(0)$). So the treatment effect is $TE_i = y_i(1) - y_i(0)$. In economic terms, the

treatment effect is the effect of participation in the NSW program on the post-program earnings of the person relative to what would have happened if the individual had not participated. More precisely: how much more (or less!) earnings would a person gain after he/she participated in the NSW job training program, relative to the situation where he/she did not participate in the program. Unfortunately, it is impossible to identify the individual treatment effect TE_i, even if we had infinite data, because by definition an individual is either treated or untreated, and we never observe the counter-factual status. However, we can identify average differences between all treated individuals (for which we observe $y_i(1)$) and untreated individuasl (for which we observe $y_i(0)$). If assignment to treatment was random, the differences between these averages gives the Average Treatment Effect $ATE = \mathbb{E}[y_i(1)] - \mathbb{E}[y_i(0)]$.

In the exercise, the outcome is the variable $re78_i$. Each individual has two potential values for this outcome $re78_i(d)$: treated status $re78_i(1)$ and $re78_i(0)$. However, we only observe the value that corresponds to the individual's treatment status. Since the treatment was randomized, the average outcome of the treated individuals is equal to the average of the population, that is $\mathbb{E}[re78_i(1)|d_i = 1] = \mathbb{E}[re78_i(1)]$ and the same holds for the untreated group: $\mathbb{E}[re78_i(0)|d_i = 0] = \mathbb{E}[re78_i(0)]$. This allows us to compute the ATE: $\mathbb{E}[re78_i(1)] - \mathbb{E}[re78_i(0)]$.

4) Questions about the data.

1. How many people are included in the dataset? 722 individuals.

2. What is the average age and level of education? 24.5 years old, and 10.3 years of education.

   Note on interpreting education years :

   - if education < 8, then junior high dropout (college in France)
   - if education < 12, then high school dropout
   - if education $= 12$, then high school completed (baccalaureat in France)
   - if education > 12, then some higher education

3. What is the proportion of the participants in the program? 41% of individuals in the dataset participated in the program.

5) Compare means of all variables for the treatment and control groups, having in mind that the NSW is a randomized program. Is the difference in the variable means as you expected?
All variables have no significant difference in their averages except for "no degree". For education, the p-value for the one-sided test suggests a rejection of the null hypothesis, butthe difference is very small: only 0.2 years difference! The outcome variable has a significant difference. These results are reassuring: given the randomization, the two groups should have identical averages, and only differ in the variables that are affected by treatment. The table below shows the p values of the two-sided test for the differences of each variable between treatment and control groups.

Table 1: Ttests

	Treated	Control	Difference	p value
age	24.45	24.63	-.18	.72
education	10.19	10.38	-.19	.14
black	.8	.8	0	.96
hispanic	.11	.09	.02	.42
married	.16	.17	-.01	.7
nodegree	.81	.73	.08	.01
re75	3026.68	3066.1	-39.42	.92
re78	5090.05	5976.35	-886.3	.06

6) Plot the empirical distribution of income in 1978, for both the treatment and the control group. [Hint: you can use a histogram *hist* or the smoothing of the histogram *kdensity*.]

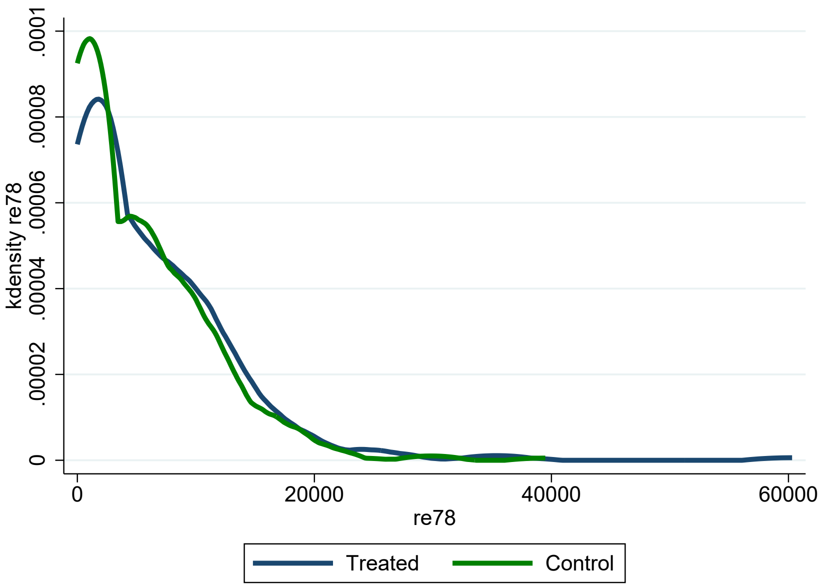


Figure 1:

Attention: We cannot identify the distribution of treatment effect (i.e. heterogeneous treatment effects) by simply looking at the distribution of the outcome variable. Only the ATE is identified without extra assumptions, even in a RCT.

7) What is the effect of the NSW on earnings? In absolute and relative terms.

The average causal effect of NSW on earnings (ATE) is the difference in the averages of the outcome variable. That is around 886 USD per year per person, that is an increase of 17.72% relative to the control group.

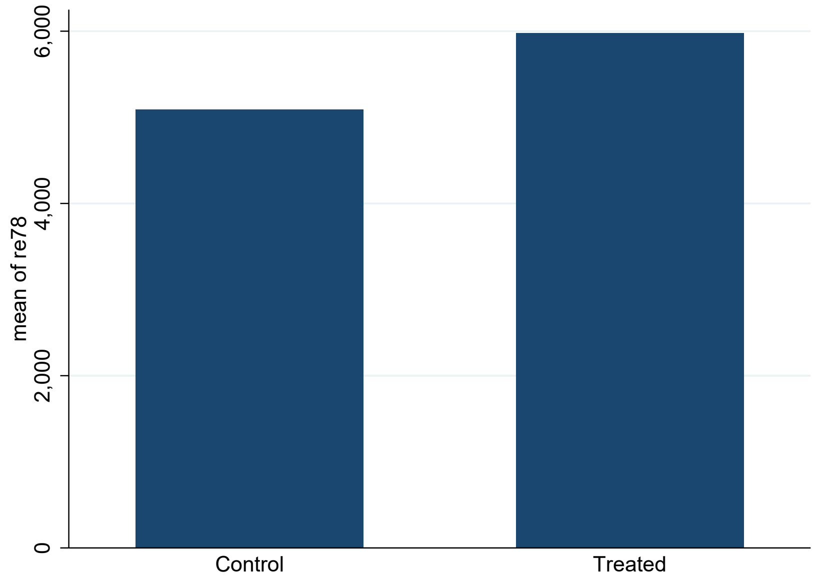


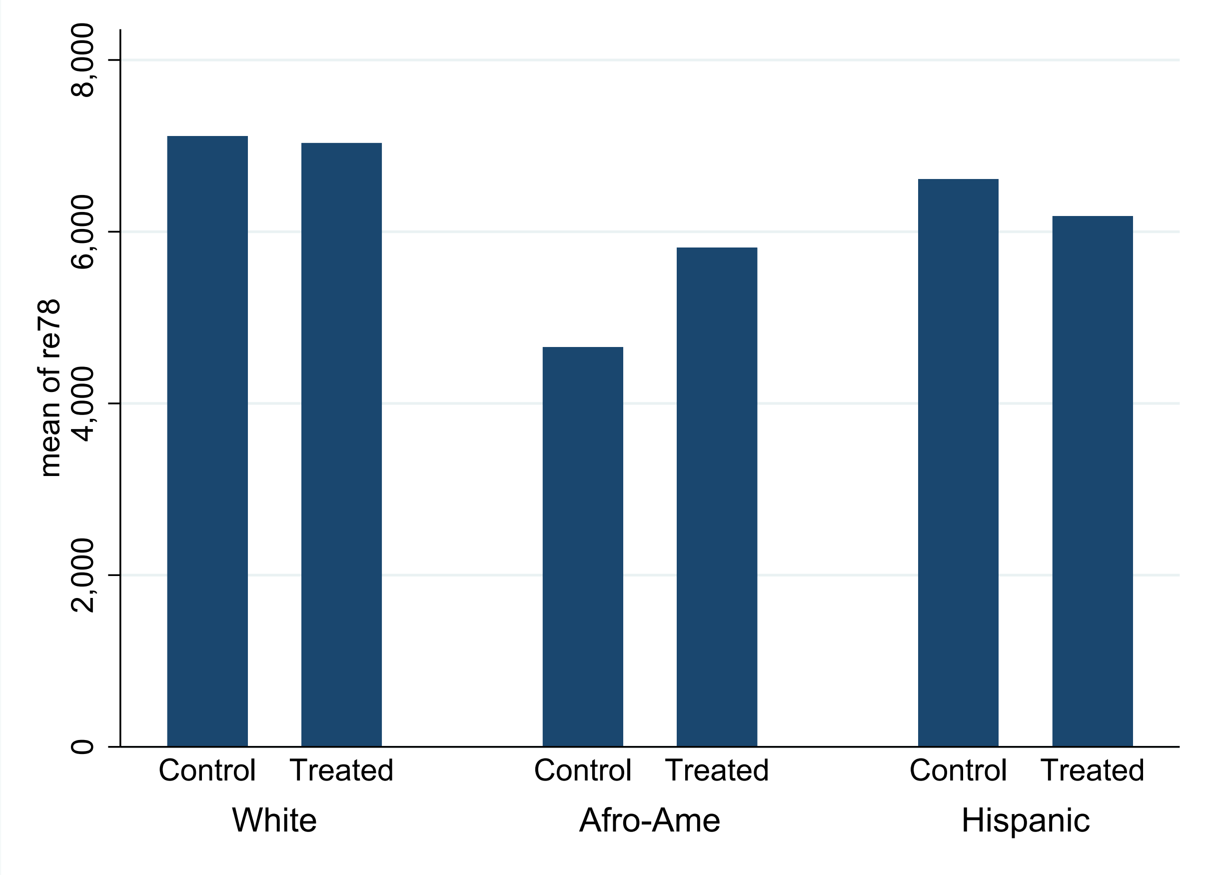
Figure 2:

8) Is the effect of the NSW on earnings significantly different from zero? How are two ways you can test this?

We reject the hypothesis of a zero-effect at the 10% confidence level, but not at the 5% confidence level. There are two simple ways to run this test. One is to run a t test comparing the means. Another way is to run a regression of the outcome on the treatment variable. $re78_i = \beta_0 + \beta_1 treat_i + \varepsilon_i$, where β_1 is the mean difference between the treatment and control group. One can then test whether β_1 is statistically different from zero.

9) Is the NSW program more effective among white, black or Hispanic people?
The program is most effective among African-Americans. There is a small negative and statistically insignificant effect on white participants and a larger negative but still statistically insignificant effect on Hispanic participants.

Figure 3: Average treatment effect of NSW on earnings, by race

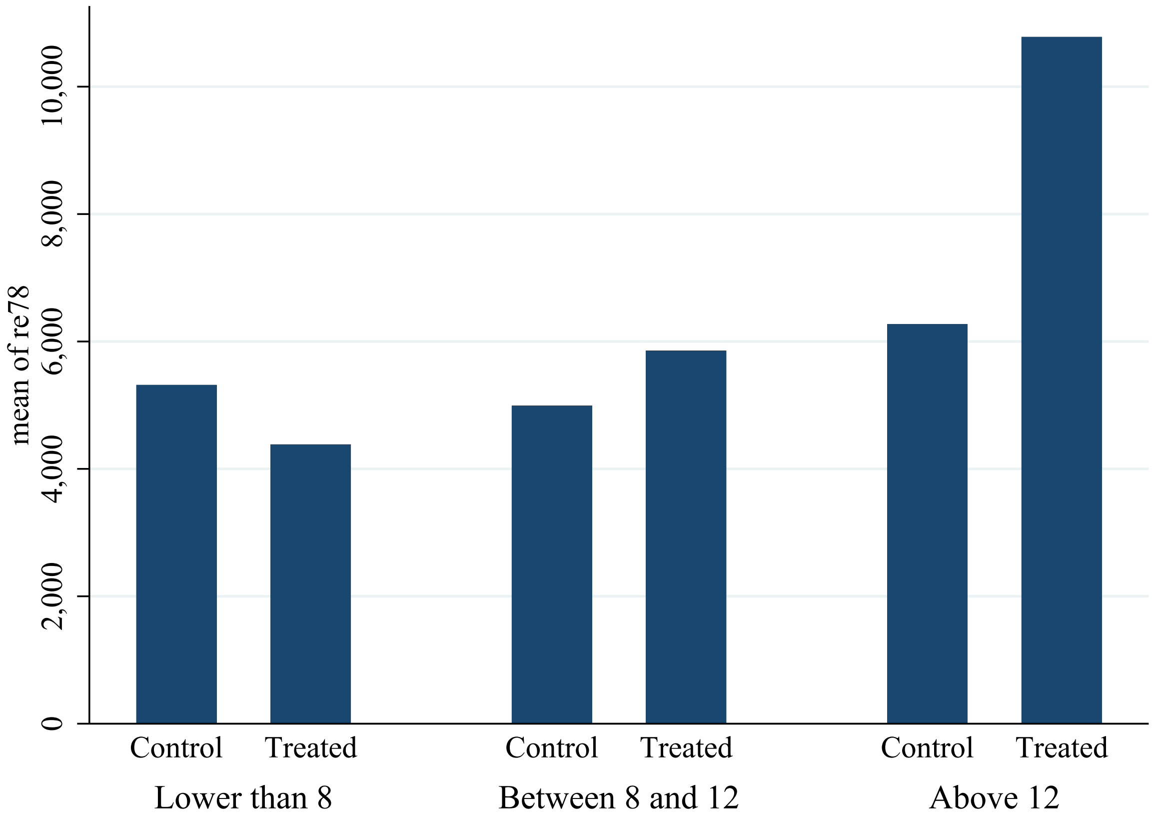


10) Is the effect of NSW homogeneous across education levels?
The estimated average difference is greatest for highest level of education, and negative for junior high school. However, the difference is only statistically significant for high school dropout (note that we obtain the same ATE as for the entire population). This is due to a too small sample size for the other subgroups, so the standard error is very high.

11) The cost of the job training was roughly 9,000 \$/head on average. In how many years the investment becomes beneficial assuming a discount factor of 0.95

Suppose that for the policymaker, the per person investment becomes profitable when the overall cost of the program per person is lower than the present discounted value of the benefits.
The average benefit per person is 886 USD in additional annual income
We are looking for the lowest number of periods T such that $cost \leq benefit(T)$

Figure 4: Average treatment effect of NSW on earnings, by groups of education level



$$
\begin{aligned}
9000 &= \sum_{t=0}^{T} 0.95^t \cdot 886 \\
\frac{9000}{886} &= \sum_{t=0}^{T} 0.95^t \\
\frac{9000}{886} &= \left(\frac{1-0.95^{T+1}}{1-0.95}\right), \quad \text{by using the formula of the geometric series} \\
\frac{9000}{886} \cdot 0.05 &= 1 - 0.95^{T+1} \\
0.95^{T+1} &= 1 - \frac{9000}{886} \cdot 0.05 \\
(T+1) \cdot \log(0.95) &= \log(1 - \frac{9000}{886} \cdot 0.05) \\
(T+1) &= \frac{\log(1 - \frac{9000}{886} \cdot 0.05)}{\log(0.95)} \\
T &= \frac{\log(1 - \frac{9000}{886} \cdot 0.05)}{\log(0.95)} - 1
\end{aligned}
$$

Re-arranging the terms, we get : $T = \frac{\log(1-\frac{9000}{886}\cdot 0.05)}{\log(0.95)} - 1 = 12.8$. That is, the program becomes profitable in a time horizon of 13 years!
The previous computation does not hold if the program has dynamic effects (i.e. if the increase in income becomes larger or fades away with time). Here we only identify the causal effect of the treatment in the first three employment years after the job training program, and assumed this increase is constant for future employment years.

12) Estimate the treatment effect, this time controlling for demographic characteristics. Does adding the controls make a difference in the estimate (and how much of a difference given the standard errors)? Why would you want to add controls even in the case where you have experimental data?

Adding controls in an RCT plays no role in terms of identification of the average treatment effect. Adding controls increases the precision of the estimation, as can be seen by the smaller standard errors. Adding the controls in our setting does not change the results much except when we control for years of education (changes the ATE to 820 USD instead of 886 USD) because of the unbalanced between the controlled and treated group in years of education. It should not do so if the randomization is successful.

Table 2: Introducing control variables

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	re78	re78	re78	re78	re78	re78
treat	886.3*	882.2*	882.8*	874.3*	869.1*	820.4*
	(472.1)	(472.3)	(470.0)	(470.6)	(470.7)	(470.2)
age		22.86	32.23	31.06	23.83	23.99
		(35.10)	(35.08)	(35.19)	(36.14)	(36.05)
black			-1,667***	-1,920**	-1,927**	-1,836**
			(581.2)	(797.4)	(797.6)	(796.9)
hispanic				-482.0	-504.1	-173.7
				(1,041)	(1,042)	(1,051)
married					569.2	517.3
					(644.9)	(643.9)
education						287.7**
						(137.7)
Constant	5,090***	4,531***	5,636***	5,921***	6,016***	2,979
	(302.8)	(910.0)	(984.0)	(1,161)	(1,166)	(1,862)
Observations	722	722	722	722	722	722
R-squared	0.005	0.005	0.017	0.017	0.018	0.024

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table created using "excel2latex" add-in for Excel:
https://www.ctan.org/tex-archive/support/excel2latex

Adding controls also allows to test for the validity of the randomization. If the randomization is successful then the controls should not have a joint significant effect, which we test using a Fisher test (and reject at the 5% level!). We have a small selection bias.

Fisher test :

R2 of the regression without controls (constrained model) : 0.0049

R2 of the regression with controls (unconstrained model) : 0.0240

$$
\begin{aligned}
F &= \frac{SSR_0 - SSR_1}{SSR_1} \times \frac{(n-k)}{q} \sim F(q, n-k) \\
&= \frac{SSR_0/SST - SSR_1/SST}{SSR_1/SST} \times \frac{(n-k)}{q} \\
&= \frac{(1-R_0^2) - 1 - R_1^2}{1-R_1^2} \times \frac{(n-k)}{q} \\
&= \frac{R_1^2 - R_0^2}{1-R_1^2} \times \frac{(n-k)}{q}
\end{aligned}
$$

14) *(Optional - Intro to Power Analysis)* Imagine to be at the experiment's design stage, when no data has been collected yet. You are to decide how many individuals to enroll in the experiment. A bigger sample decreases the minimum detectable effect, but is also more expensive to obtain. The power of a test is the probability of correctly rejecting the null hypothesis when the null hypothesis is false. That is, $\pi = 1 - \beta = \Pr(\text{reject H0}|\text{H0 is false})$. Therefore: the higher the power, the larger the sample size required; The larger the effect size, the smaller the sample size; the larger

the significance level, the smaller the sample size, with everything else being equal. Using 1975 income (*re75*) to benchmark your study:

Let δ be the difference in the (continuous) outcome variable between the treatment and the control group, with respective size of n_1 and n_0, and assume that the two groups have the same estimated variance $\hat{\sigma}^2$. The minimum detectable effect at a power of $\pi = 1 - \beta$ is simply:

$$\delta = (t_\beta + t_{\alpha/2})\hat{\sigma}^2 \sqrt{\frac{1}{n_1} + \frac{1}{n_0}} \tag{1}$$

where t_x is the expression for the t-distribution. Stata has a built-in command to calculate (or approximate) this formula, fixing power, sample size or minimum detectable effect.