

**Как да осигурим постфактум сходни двойки (matching cases) при
оценка на нетния ефект на обучения**
**Доц. д-р Калоян Валентинов Харалампиев, Софийски университет
„Св. Климент Охридски“, Философски факултет, катедра
„Социология“**

**Йордан Димитров Димитров, Балкански институт по труда и
социалната политика**

**How to Provide Afterwards Matching Cases in Assessing the Net Effect of
Trainings**

**Associate professor Kaloyan Valentinov Haralampiev PhD, Sofia University
“St. Kliment Ohridski”, Faculty of Philosophy, Department of Sociology
Yordan Dimitrov Dimitrov, Balkan Institute for Labour and Social Policy**

Резюме: Проблемната ситуация има няколко важни особености:

- (1) Целта е да се оцени ефектът на образователни/обучителни програми. Това може да са чуждоезикови обучения, компютърни обучения, програми за квалификация и преквалификация. Особеността е, че съвкупността от лица, които подлежат на такова образование/обучение, е достатъчно голяма, но участието е доброволно. Това прави скъпо провеждането на входно изследване, тъй като предварително не е ясно кой ще участва в обучението и кой не, което означава, че входното изследване трябва да обхване цялата съвкупност, а тя е голяма.
- (2) Затова вместо да се прави и входно, и изходно изследване, може да се направи само изходно изследване, като експерименталната група е извадка от лицата, участвали в обучението.
- (3) За да се минимизира, доколкото е възможно, влиянието на страничните фактори, е добре контролната група да бъде формирана на принципа на сходните двойки (matching cases), т.е. на всяко лице от експерименталната група да бъде съпоставено лице от контролната група, идентично по набор от признаци с лицето от експерименталната група, но което не е взело участие в обучението. На теория това е лесно да се направи, но на практика винаги се получават откази както от експерименталната, така и от контролната група.

Това означава, че във вече събраните данни, част от лицата формират сходни двойки, но за част от лицата от експерименталната група няма аналог в контролната група, както и обратното – за част от лицата в контролната група няма аналог в експерименталната група. Идентифицирането на сходните двойки и на лицата, за които няма аналог, е достатъчно трудоемка работа, затова се търсят решения за нейното минимизиране. Тривиалното решение е в този случай да се използва логистична регресия. В нашия доклад ние предлагаме нетривиално решение чрез претегляне с подходяща система от тегла.

Емпиричните данни на нашето изследване са от реално проведена оценка на ефекта на програми за преквалификация на безработни.

Ключови думи: обучения, нетен ефект, квази-експериментален дизайн, сходни двойки, логистична регресия, претегляне

Abstract: The problem has several important characteristics:

- (1) The aim is to assess the effect of education/training programs. These may be foreign language courses, computer training, programs for training and retraining. The peculiarity is that the population of persons, who are subject to such education/training, is large enough, but the participation is voluntary. The latter makes any pre-tests quite expensive, as initially it is undefined who will participate in the training and who will not, which means that the pre-test should cover the whole population, and it is large enough.
- (2) Instead of conducting both pre-test and post-test it is rather having only post-test where the experimental group is a sample of those who participated in the training.
- (3) In order to minimize as far as possible the impact of extraneous factors, the control group is preferably to be formed on the principle of the matching cases. This means any person from the experimental group should be compared to a person from the control one who is identical by set of

characteristics, but has not participated in any training. Theoretically, it is easy to be implemented, but in practice there are always refusal respondents in both groups.

Thus in the already collected data some persons form matching cases, but others in the experimental group have no analogue in the control group, and vice versa – there are individuals in the control group who have no analogue in the experimental one. Identification of matching cases and such with no analogue is a labor-intensive work that needs solutions for minimization. A trivial solution is the logistic regression. But in this paper, a non-trivial solution is suggested by using an appropriate system of weights and respective approach of weighing.

Empirical data of our study are from existent assessment of the effect of retraining programs for unemployed people.

Keywords: training, net effect, quasi-experimental design, matching cases, logistic regression, weighting

При провеждане на обучения¹ целта е да се оцени т.нар. нетен ефект на образователни/обучителни програми. Под нетен ефект се разбира чистият ефект от участието в програмата, като е изчистено влиянието на всички останали възможни фактори. За да може да стане това, трябва дизайнът на изследването да бъде експериментален (Уолш, Коцева, Доле, Доренбос, 2001, 21). На практика обаче участието в тези програми е доброволно, което означава, че е невъзможно експерименталната и контролната група да бъдат формирани по случаен начин. Затова практическото решение е дизайнът на изследването да бъде квази-експериментален (Уолш, Коцева, Доле, Доренбос, 2001, 21-22). Но проблемът се поражда от това, че няма как предварително да се знае кой доброволно ще се включи в обучението и кой не. В такъв случай едно входно изследване (pre-test) би било доста скъпо, защото трябва да обхване цялата съвкупност, а тя е голяма. Затова вместо да се прави и входно, и изходно изследване, може да се направи само изходно изследване (post-test), като експерименталната група е извадка от лицата, участвали в обучението.

В нашия случай за оценка на нетния ефект от активни програми и мерки за заетост и обучение на безработни беше формирана извадка от 913 младежи, участвали в осем различни програми, като всяко лице е участвало само в една програма. На тази експериментална група беше съпоставена контролна група също от 913 изследвани лица.

За да се минимизира, доколкото е възможно, влиянието на страничните фактори, е добре контролната група да бъде формирана на принципа на сходните двойки (matching cases), т.е. на всяко лице от експерименталната група да бъде съпоставено лице от контролната група, идентично по набор от признаци с лицето от експерименталната група, но което не е взело участие в обучението (Уолш, Коцева, Доле, Доренбос, 2001, 29).

Извадката от лицата в експерименталната група беше стратифицирана по местожителство и по конкретната програма, а подборът във всяка страта беше извършен с използването на стъпка. Извадката на лицата в контролната група също беше стратифицирана по местожителство, но още и по пол, възраст и образование. Анкетите получиха инструкция, според която за всеки анкетиран младеж, участвал в програма, трябваше да намерят младеж, неучаствал в програми, със същия пол, възраст и образование. Възрастта беше разделена на три групи (15-19, 20-24 и 25-29 години), а образованието – на шест групи (висше, полувисше, средно специално, средно, основно и начално и по-ниско). По този начин планираният обем на извадката от младежи, неучаствали в програми, е същият като планирания обем на извадката от младежи, участвали в програми, и структурата на двете извадки по пол, възраст и образование е идентична.

¹ Това може да са чуждозикови обучения, компютърни обучения, програми за квалификация и преквалификация и т.н..

На теория това е лесно да се направи, но на практика винаги се получават откази както от експерименталната, така и от контролната група.

В нашия случай след приключване на работата на терен получихме попълнени 627 анкетни карти от лицата в експерименталната група. Това представлява 68,7% от планираната извадка. От лицата в контролната група получихме попълнени 562 анкетни карти, което представлява 61,6% от планираната извадка.

Това означава, че във вече събраните данни, част от лицата формират сходни двойки, но за част от лицата от експерименталната група няма аналог в контролната група, както и обратното – за част от лицата в контролната група няма аналог в експерименталната група. Затова сравнихме структурата на двете групи по пол, възраст и образование.

Таблица 1
Структура на двете групи по пол (%)

Пол	Експериментална група	Контролна група
Жена	56,6	56,4
Мъж	43,4	43,6
Общо	100,0	100,0

Наблюдаваните разлики в процентите са статически незначими².

Таблица 2
Структура на двете групи по образование (%)

Образование	Експериментална група	Контролна група
Без образование	1,3	0,5
Начално	1,0	1,6
Основно	5,4	5,0
Средно гимназиално	17,8	21,4
Средно специално	31,5	31,8
Полувисше	0,8	0,2
Висше	42,2	39,5
Общо	100,0	100,0

Наблюдаваните разлики в процентите са статически незначими.

Средната възраст на младежите в експерименталната група е 27,0 години, а средната възраст на контролната група е 26,5 години. Разликата от половин година е статистически значима³, но е достатъчно малка, за да очакваме, че би оказала изкривяващо влияние върху направените от нас изводи. Основанието ни да я считаме за достатъчно малка е, че стандартните отклонения в двете групи са 2,8 години, т.е. различията вътре в двете групи са многократно (повече от пет пъти) по-големи отколкото различията между двете групи. Коефициентът на определеност е 0,010, което означава, че само 1,0% от различията във възрастта се обясняват от принадлежността към различните групи.

² За сравняване на две разпределения е използван Хи-квадрат. Разликите се приемат за значими при равнище на значимост по-малко от 0,05.

³ За сравняване на две средни е използван t-тест. Разликите се приемат за значими при равнище на значимост по-малко от 0,05.

В резултат от направените проверки можем да твърдим, че структурата на експерименталната и контролната група е идентична и можем да сравняваме резултатите, получени от двете извадки.

Но нетният ефект се търси не общо, а поотделно за всяка програма. Тоест експерименталната група не е една, а са осем. И на всяка една експериментална група трябва да се съпостави отделна част на контролната група по принципа на сходните двойки.⁴

Идентифицирането на сходните двойки и на лицата, за които няма аналог, е достатъчно трудоемка работа, затова се търсят решения за нейното минимизиране.

Тривиалното решение в този случай е да се използва логистична регресия (Уолш, Коцева, Доле, Доренбос, 2001, 29). Това е така, защото коефициентите на логистичната регресия показват чистото влияние на всеки фактор, при условие че останалите фактори не се променят (Чипева, 2005, 91). Тоест, ако в нашия модел зависимата променлива е намирането на работа, а независимите променливи са участието в конкретна програма за заетост и обучение, пол, възраст и образование, то коефициентът на участието в конкретна програма ще покаже какво е чистото влияние на участието в тази програма, при условие че се сравняват лица с един и същи пол, възраст и образование.

Ние предлагаме нетривиално решение чрез претегляне с подходяща система от тегла.

Най-лесният начин да се получи изравняване на структурите на експерименталната и контролната група е при сравняването на контролната група с експериментална група от младежи, участвали в конкретна програма, за контролната група да се използват тегла, които се изчисляват по формулата:

$$w_{ij} = \frac{\frac{n_{ij}}{n_j}}{\frac{n_{ik}}{n_k}},$$

където w_{ij} е теглото в стратата⁵ i в контролната група при сравняването ѝ с програма j , $\frac{n_{ij}}{n_j}$ е дялът на стратата i в програмата j , а $\frac{n_{ik}}{n_k}$ е дялът на стратата i в контролната група.

По този начин, използвайки тегла, структурата на контролната група по пол, възраст и образование се изравнява със структурата на конкретната експериментална група, с която се сравнява. В резултат на това можем да получим ажустирани оценки на нетния ефект.

Ще илюстрираме този подход с един пример. Ние сравнявахме експерименталната и контролната група по 12 показателя, но най-важният от тях е дали безработните са си намерили работа.

При използването на логистична регресия със зависима променлива намирането на работа и независими променливи участието в конкретна програма, пол, възраст и образование, установихме, че само Програма №5 и Програма №8 оказват значимо

⁴ Би могло всяка една от осемте експериментални групи да се сравнява с цялата контролна група, но тогава се получават т.нар. неажустирани оценки (Уолш, Коцева, Доле, Доренбос, 2001, 28), при които не може да каже каква част от различието в процента на намерили работа се обяснява от участието в конкретната програма и каква част – от други възможни фактори.

⁵ Стратите са формирани по пол, възраст (три възрастови групи) и образование (шест групи по образование).

влияние върху намирането на работа. За съжаление влиянието е отрицателно – участието в тези две програми намалява шанса за намиране на работа на свободния пазар, при условие че полът, възрастта и образованието са константни.

Таблица 3

Резултати от логистичната регресия със зависима променлива намиране на работа на свободния пазар (референтна група – не си намерих работа на свободния пазар)

Фактори		B	Sig.	Exp(B)
Програма (референтна група – контролна група)	Програма 1	-0,039	0,898	0,962
	Програма 2	-0,008	0,987	0,992
	Програма 3	-0,610	0,257	0,544
	Програма 4	-0,424	0,126	0,655
	Програма 5	-1,345	0,006	0,261
	Програма 6	-0,401	0,227	0,670
	Програма 7	-0,020	0,914	0,980
	Програма 8	-1,063	0,004	0,345
Пол (референтна група – мъж)	Жена	0,020	0,885	1,020
Образование (референтна група – полувисше и висше)	Основно и по-ниско	-1,339	0,000	0,262
	Средно гимназиално	-0,608	0,002	0,544
	Средно специално	-0,659	0,000	0,517
Възраст		0,020	0,433	1,021
Константа		0,160	0,828	1,173

Използването на тегла потвърди този извод.

Таблица 4

Разпределение на изследваните лица по намирането на работа на свободния пазар (%)

	Програма 1	Контролна група	Общо
Не си намерих работа на свободния пазар	33,3	35,8	35,6
Намерих си работа на свободния пазар	66,7	64,2	64,4
Общо	100,0	100,0	100,0

Наблюдаваните разлики в процентите са статически незначими.

Таблица 5

Разпределение на изследваните лица по намирането на работа на свободния пазар (%)

	Програма 2	Контролна група	Общо
Не си намерих работа на свободния пазар	62,5	66,2	66,0
Намерих си работа на свободния пазар	37,5	33,8	34,0
Общо	100,0	100,0	100,0

Наблюдаваните разлики в процентите са статически незначими.

Таблица 6

Разпределение на изследваните лица по намирането на работа на свободния пазар (%)

	Програма 3	Контролна група	Общо
Не си намерих работа на свободния пазар	56,3	39,1	39,6
Намерих си работа на свободния пазар	43,8	60,9	60,4
Общо	100,0	100,0	100,0

Наблюдаваните разлики в процентите са статически незначими.

Таблица 7

Разпределение на изследваните лица по намирането на работа на свободния пазар (%)

	Програма 4	Контролна група	Общо
Не си намерих работа на свободния пазар	59,1	50,6	51,6
Намерих си работа на свободния пазар	40,9	49,4	48,4
Общо	100,0	100,0	100,0

Наблюдаваните разлики в процентите са статически незначими.

Таблица 8

Разпределение на изследваните лица по намирането на работа на свободния пазар (%)

	Програма 5	Контролна група	Общо
Не си намерих работа на свободния пазар	76,0	53,5	54,6
Намерих си работа на свободния пазар	24,0	46,5	45,4
Общо	100,0	100,0	100,0

Наблюдаваните разлики в процентите са статически значими.

Таблица 9

Разпределение на изследваните лица по намирането на работа на свободния пазар (%)

	Програма 6	Контролна група	Общо
Не си намерих работа на свободния пазар	53,7	44,6	45,2
Намерих си работа на свободния пазар	46,3	55,4	54,8
Общо	100,0	100,0	100,0

Наблюдаваните разлики в процентите са статически незначими.

Таблица 10

Разпределение на изследваните лица по намирането на работа на свободния пазар (%)

	Програма 7	Контролна група	Общо
Не си намерих работа на свободния пазар	42,2	39,1	39,9
Намерих си работа на свободния пазар	57,8	60,9	60,1
Общо	100,0	100,0	100,0

Наблюдаваните разлики в процентите са статически незначими.

Таблица 11

Разпределение на изследваните лица по намирането на работа на свободния пазар (%)

	Програма 8	Контролна група	Общо
Не си намерих работа на свободния пазар	69,2	41,4	43,3
Намерих си работа на свободния пазар	30,8	58,6	56,7
Общо	100,0	100,0	100,0

Наблюдаваните разлики в процентите са статически значими.

Важно е да се отбележи, че процентите в колоната „Контролна група“ са различни във всичките осем таблици. Това е така, защото контролната група е винаги различна, т.е. с използването на тегла ние сме конструирали осем различни контролни групи, идентични по структура с осемте различни експериментални групи.

Предимствата на използването на тегла за контролната група вместо формиране на сходни двойки от изследвани лица са две:

- Използването на тегла е много по-лесно, отколкото формиране на сходни двойки;
- При използването на тегла не се губи информация, защото всички изследвани лица от контролната група участват в сравнението, само че получават различна тежест. При използването на сходни двойки част от изследваните лица отпадат при сравнението и информацията за тях се губи.

Библиографска справка

Уолш, К., М. Коцева, Е. Доле, Р. Доренбос. (2001). Оценка на нетното въздействие на активните програми на пазара на труда в България. София: Министерство на труда и социалната политика

Чипева, С. (2005). Статистически анализ на категорийни данни с SPSS. София: Университетско издателство „Стопанство“

Калоян Валентинов Харалампиев k_haralampiev@phls.uni-sofia.bg;

Йордан Димитров Димитров jordandim@bilsp.org