ECONOMICS 705
EVALUATION OF SOCIAL PROGRAMS

Prof. Jeffrey Smith
Department of Economics
University of Wisconsin-Madison
Fall 2022

PROBLEM SET 1
ANSWERS
Version of September 26, 2022

1. (5 points) See the Stata log file.

2. (5 points) See the Stata log file.

3. (5 points) See the Stata log file.

4. (5 points) The differences are small in both absolute and relative (i.e. percentage) terms. For example, the difference for earnings last year equals about $42.30, or roughly 2.2 percent of the control group mean.

5. (5 points) Perhaps surprisingly, the small estimated difference in years of average years of schooling in the treatment and control groups of about 0.16 years or 1.4 percent of the control mean differs statistically from zero with a p-value of 0.0145. The differences for age and earnings last year do not come close to statistical significance.

Properly implemented random assignment assures the truth of the null hypothesis in the population; in that sense, as noted by Deaton and Cartwright (2018), it represents an odd null to test. What the researcher really cares about is covariate balance in the sample. In the JTPA Study context, neither the test results nor the institutional background suggest any skepticism about the implementation of random assignment.

6. (5 points) The standardized difference is

$$SDIFF = 100\frac{|\bar{X}_{R=1} - \bar{X}_{R=0}|}{\sqrt{\dfrac{\widehat{\text{var}}(X)_{R=1} + \widehat{\text{var}}(X)_{R=0}}{2}}}$$

Plugging in the numbers in this context yields

$$SDIFF \approx 100 \frac{|11.46 - 11.30|}{\sqrt{\frac{(2.25)^2 + (2.41)^2}{2}}} = 100 \frac{0.16}{\sqrt{\frac{5.0625 + 5.8081}{2}}}$$

$$= 100 \frac{0.16}{2.3314} = (100)(0.0686) \approx 6.86$$

This standardized difference is modest but not trivial by the standards of the literature. In particular, it is about one-third of the arbitrary (though not meaningless and perhaps a bit large) standard of 20 suggested by Rosenbaum and Rubin (1985).

7. (5 points) The estimated mean impact on earnings in the 18 months after random assignment equals \$792.81 with a standard error of about \$228.53. The p-value from the two-sided test of the null that the population mean difference in earnings in the 18 months after random assignment equals zero is 0.0005. The difference is thus strongly statistically significant, as the p-value is less even than 0.001, let along 0.01 or 0.05. The impact is more than 10 percent of the control mean, and so substantively meaningful, though not huge.

Note the negative sign on the mean impact estimate reported by Stata in this case. Stata takes the control group mean minus the treatment group mean rather than the reverse, which does not affect the results of the t-test but does affect the meaning of the impact estimate!

8. (5 points) As expected (because the mathematics force it to be so), calculating the mean difference experimental impact estimate using a simple linear regression with *esum18i* as the dependent variable and the treatment indicator as the independent variable yields exactly the same estimate as using the `ttest` command. The standard error, equal to \$223.52 differs slightly, as do the t-statistic and p-value, because of the use of the "comma robust" variance estimator. The regular OLS standard errors, and their associated t-statistic and p-value, would exactly match those produced by the `ttest` command.

9. (5 points) The mean impact estimate in this case equals about \$694.61 with a (robust) standard error of about \$209.40. The t-statistic equals 3.32 and the p-value equals 0.001, so that the estimate is again strongly statistically significant.

As expected, including baseline covariates in the regression reduces the estimated (robust) standard error (from around \$223.52 to around \$209.40), though not by all that much. The baseline covariates reduce the variance of the outcome equation residual but are not, by construction, correlated with the treatment indicator.

Due to random assignment, both the estimator in this problem and the estimator in Problem 6 are unbiased. Recall that unbiasedness is a property of the estimator, i.e. of the process or rule used to generate the estimate. In a given finite sample, two unbiased estimators will typically yield similar, but not identical, estimates, as they do in this case.

10. (5 points) Around 65.89 percent of the treatment group enrolls in JTPA in the National JTPA Study data, compared to only about 2.66 percent of the control group.

11. (5 points) The IV estimate obtained via two-stage least squares estimation equals about \$1253.89, with a standard error of \$353.71. As expected, it substantially exceeds the simple mean difference estimate in magnitude because it implicitly assumes that those who do not enroll in JTPA have a treatment effect of zero. Like the simple mean difference estimate, it is both substantively and statistically significant, with a p-value from the t-test of the null of a zero population coefficient equal to 0.000.

In a homogeneous effects world the IV estimate represents the common effect of JTPA enrollment. In a heterogeneous effects world the IV estimate represents the Local Average Treatment Effect (LATE), which is the average treatment effect on those who enroll in JTPA when in the treatment group but not when in the control group. As noted in class, the literature calls this group the compliers. The LATE interpretation requires a monotonicity assumption, i.e. that there are no defiers who would enroll in JTPA if assigned to the control group but not if assigned to the treatment group.

12. (5 points) To obtain an estimate of the mean education among the compliers, recall that, under monotonicity, we have only three groups: always-takers, compliers, and never-takers. The treatment group members who do not receive any services provide an estimate of the fraction of never-takers in the population as well as an estimate of their population mean years of schooling. In the data, these equal (1.0000 – 0.6589) = 0.3411 and 11.40, respectively. Similarly, the members of the experimental control group who receive services provide an estimate of the fraction of always-takers in the population as well as an estimate of their population mean years of schooling. These estimates equal 0.0266 and 11.29, respectively.

Under monotonicity, the unconditional mean of *bfeduca* (abbreviated as *bfed* in the equation) constitutes a weighted average of the means for the always-takers, the compliers, and the never-takers. In notation, we have

$$E(bfed) = \Pr(AT)E(bfed \mid AT) + \Pr(C)E(bfed \mid C) + \Pr(NT)E(bfed \mid NT)$$

Noting that $\Pr(C) = 1 - \Pr(AT) - \Pr(NT)$ and plugging in the sample analogue estimates from the data yields

$$(11.41) = (0.0266)(11.29) + (1 - 0.0266 - 0.3411)\hat{E}(bfed \mid C) + (0.3411)(11.40)$$

Rearranging a bit yields

$$\hat{E}(bfed \mid C) = \frac{(11.41) - (0.0266)(11.29) - (0.3411)(11.40)}{(1 - 0.0266 - 0.3411)} = \frac{11.41 - 0.3003 - 3.8885}{0.6323}$$

Solving gives an estimate of 11.42. Substantively, the always-takers, the never-takers, and the compliers all have very similar values for average years of schooling.

13. (5 points) See the Stata log file.

14. (5 points) The subgroup estimates (and associated standard errors) obtained from the separate regressions are \$671.83 (290.67) for ages 31 and below and \$705.94 (301.38) for ages 32 and above. Thus, the estimate for the older population is about \$30 more than that for the younger population. As expected, the pooled estimate given above lies in between the two subgroup estimates.

As the subgroup estimates obtained in this way are statistically independent (and so have zero covariance) one can test the null of equal population impacts by looking at whether one estimate lies in the confidence interval of the other. Not surprisingly given the small substantive difference and reasonably large sample sizes, this is indeed the case, meaning that the data do not reject the null of equal mean impacts for older and younger participants (as well as many other nulls involving non-zero differences).

The estimated impact for ages 32 and above obtained from the single regression with the interaction term equals \$715.86 (320.18). The estimate for ages 31 and below equals (715.86 – 44.15) = \$671.71. In this setup, a test that the population coefficient on the interaction term equals zero corresponds to the null that the impact for older participants equals that for younger participants. Clearly, the data do not reject that null in the pooled specification given the p-value of 0.916.

Two opposing forces affect the standard errors. First, estimating two separate models means smaller effective sample sizes, which acts to increase the standard errors. At the same time, if the population coefficients on the conditioning variables differ substantially between older and younger individuals, allowing them to differ in the estimation will reduce the residual variance for both groups, and so act to reduce the standard errors. In this case, the net is a slightly smaller standard error on the treatment effect for older participants in the pooled model.

Full credit here depends on recognizing the tradeoffs and making a case one way or the other, rather than choosing a specific option.

15. (5 points) The standard deviation of earnings in the 18 months after random assignment in the treatment group is about \$8239.86 while that in the control group is about \$7738.12. The standard deviation for the control group exceeds the control group mean, while the standard deviation for the treatment group roughly equals its mean. Earnings in low-skill populations have a high variance in general, in part because of all the zeros.

The p-value from the statistical test of the null of equal variances performed by the `sdtest` command equals 0.0019 indicating strong statistical evidence of a difference in population variances. The treatment group variance exceeding the control group variance corresponds to what we would expect if the heterogeneous treatment effects add a variance component to the treatment group outcomes not present in the control group outcomes (and that component does not have a negative correlation with the untreated outcome).

16. (5 points) The quantile treatment effect at the 75th percentile of earnings in the 18 months after random assignment equals \$1007.00 with a standard error of about \$396. It exceeds the average treatment effect by about \$200.

17. (5 points) With no assumptions other than the validity of random assignment, the quantile effects can be interpreted as the impact of treatment *on* the particular quantile of the outcome distribution. That is, one can make statements such as "the estimated impact of treatment on the 75th percentile of the outcome distribution equals about \$1007. With the additional assumption of rank preservation, so that a given quantile of the treated distribution represents the counterfactual for the same quantile of the untreated outcome distribution and vice versa, we can interpret the impact estimates as impacts *at* (rather than *on*) particular quantiles of the outcome distribution. That is, we can make statements such as "the estimated impact of treatment on individuals at the 75th percentile of the outcome distribution equals about \$1007."

18. (5 points) The probability of employment (i.e. having positive earnings in the 18 months after random assignment) equals 0.7954 in the treatment group and 0.7652 in the control group.

19. (5 points) The formula for the bounds on the probability of (0, 0) cell in the 2 x 2 case is given by:

$$\max\{(p_{00}+p_{10})+(p_{00}+p_{01})-1,0\} \leq p_{00} \leq \min\{p_{00}+p_{01}, p_{00}+p_{10}\}.$$

Given the values in the preceding problem, we have the following:

$$(p_{00}+p_{10}) = 1.0000-0.7652 = 0.2348$$
$$(p_{00}+p_{01}) = 1.0000-0.7954 = 0.2046$$

Thus, the bounds on the (0,0) cell probabilities are given by:

$$\max\{0.2348+0.2046-1.0000, 0.0000\} = 0.0000 \leq p_{00} \leq \min\{0.2046, 0.2348\} = 0.2046 \approx 0.20$$

In words, given the observed marginal distributions, the probability of the cell corresponding to zero earnings in the 18 months after random assignment in both the treated and control states lies between 0.0000 and 0.2046, inclusive.

The argument that these bounds are narrow is that they eliminate a very large range of possibilities, namely [0.2046, 1.0000]. At the same time, the argument that they are wide is that a substantial range of possible cell probabilities consistent with the given marginals remains.

20. (5 points) At the lower bound distribution, the four cell probabilities become

$$p_{00} = 0.0000$$
$$p_{01} = 0.2348$$
$$p_{10} = 0.2046$$
$$p_{11} = 1 - (0.2348 + 0.2046) = 0.5606$$

The second and third value follow immediately from the marginals. The final value follows from the first three plus the fact that the four cell probabilities sum to 1.0000.

The observations on the diagonal cells have treatment effects equal to zero because they have the same (binary) outcome in the treatment and control states. Thus, 0.2046 have a treatment effect of -1.0, 0.5606 have a treatment effect of 0.0, and 0.2348 have a treatment effect of 1.0.

At the upper bound distribution,

$$p_{00} = 0.2046$$
$$p_{01} = 0.2348 - 0.2046 = 0.0302$$
$$p_{10} = 0.0000$$
$$p_{11} = 0.7652$$

In this case, (0.2046 + 0.7652) = 0.9698 have a treatment effect of 0.0 while the remainder, 0.0302, have a treatment effect of 1.0.

As with the example in class, these are very different treatment effect distributions indeed, and yet both are completely consistent with the given marginals.