

WorkFirst Activities:

To get impact on Earnings:

- 1) **Run OLS regression** to get coefficients showing effects of each activity (and controls). These compare those who completed an activity with those who did not.
- 2) **Calculate predicted earnings for those who DID complete an activity two ways:**

With dummy for activity=1: $\hat{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_m X_m$

With dummy for activity =0: $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_m X_m$

[Note: You can use Stata to do this with a PREDICT command by creating a new variable for \hat{Y}_0 and one for \hat{Y}_1 . You do this separately for each activity.]

- 3) **Take the average of \hat{Y}_0** to get our prediction of what earnings would be for those who completed the activity IF they had not had it (our counterfactual). This reflects their other characteristics (including any other activities they completed).
- 4) Our **estimate of the effect of the program** is the difference between the two predictions $\hat{Y}_1 - \hat{Y}_0$. In this linear model, it will always be $\hat{\beta}_1$.
- 5) **Graph predicted probability** without activity on bar \hat{Y}_0 and difference between that and probability of employment WITH activity $\hat{\beta}_1$.

To get impact on Employment:

- 1) **Run logit model** to get coefficients showing effects of each activity (and of other control variables). This again compares those who got the activity with those who did not.
- 2) **Calculate probabilities** with each work activity =1 and =0 for each case that had (work activity =1) using SPSS.

Probabilities from logit model:

$$\hat{P} = \frac{1}{1 + e^{-Z}} \text{ where } Z = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_m X_m$$

In Stata again use the PREDICT command first using REPLACE X=1 (had the activity) then REPLACE X=0 (didn't have the activity) for each type of training (separately).

- 3) **Get average of predicted probabilities** with and without activity for **subsample that actually completed each activity** (separately) So in Stata SUMMARIZE Yhat1 and YHAT2 (predicted value with and without activity) for a subset of observations (IF X==1)
- 4) **Graph predicted probability** without activity on bar \hat{P}_0 and difference between that and probability of employment WITH activity $\hat{P}_1 - \hat{P}_0$. Note that here the difference is not just $\hat{\beta}_1$ --the addition to the probability of employment varies depending on Z.