Peter Faber
Regression Analysis Student Project

Goal

The goal of this analysis is to examine the effectiveness of a mid-1970s job training program on low-income workers who lacked basic labor skills to enter the job market. The National Supported Work Demonstration was a program designed to help hard-to-employ people get and hold, normal unsubsidized jobs. It was offered at 15 locations around the country and lasted 12 to 18 months. A regression analysis will be completed to determine if participation in the NSWD was effective in helping hard-to-employ people get jobs.

Data

The data used in this analysis was found from the ICPSR website and slightly modified for the Brown University Urban Education Policy master's program. It was collected from a randomized experiment. The original data can be found via the Interuniversity Consortium for Political and Social Research Website at http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/7865. The data used in this analysis, which omits some group members who did not report earnings in 1974, is attached in the file titled "PF NSWD Regression Analysis.xlsx" on the 'nsw' tab.

The following provides a brief description of the data:

Data	Description
nsw	A dummy variable where 1 indicates that a person participated and 0 indicates that they did not.
age	The age of the subject at the time of the study.
educ	The number of years of education each subject received.
married	A dummy variable where 1 indicates that a person is married and 0 indicates that they were not.
earn_75	January 1, 1975 – December 31, 1975 earnings
earn_78	January 1, 1978 – December 31, 1978 earnings
noncompleter	A dummy variable where 1 indicates that a person did not complete the NSWD and 0 indicates that they did.

The following provides a summary of the potential explanatory variables relating to the effectiveness of the NSWD for the entire population:

Variable	Number of Observations	Mean	Std. Dev.	Min	Max
Age	578	24.829	6.734	17	55
Education	578	10.325	1.661	3	16
Marital Status	578	0.164	0.371	0	1
1975 Earnings	578	2,831.23	4,765.89	0.00	37,431.66

We consider the same measurements separated for the group which participated in the NSWD (nsw=1) and the group which did not participate in the NSWD (nsw=0).

Participated					
Variable	Number of Observations	Mean	Std. Dev.	Min	Max
Age	238	24.853	6.733	17	48
Education	238	10.382	1.849	4	16
Marital Status	238	0.168	0.375	0	1
1975 Earnings	238	2,785.92	4,756.44	0.00	37,431.66

Did Not Participate					
Variable	Number of Observations	Mean	Std. Dev.	Min	Max
Age	340	24.812	6.745	17	55
Education	340	10.285	1.518	3	14
Marital Status	340	0.162	0.369	0	1
1975 Earnings	340	2,862.95	4,779.24	0.00	32,984.25

Comparing the basic measures of group of subjects who participated in the NSWD and those subjects who did not demonstrates that the groups are well balanced. There are subtle differences in the data but they are relatively small when compared to the magnitude of what is being measured.

Analysis

All analysis was completed using the Microsoft Excel Regression Toolpak. The attached spreadsheet, "PF NSWD Regression Analysis.xlsx" contains the raw output as created by Excel on the 'results' tab. Throughout the analysis, effectiveness of the program will be measured by 1978 salary. It is assumed that subjects with high 1978 salary (salary after the program was complete) are 'successful'.

We consider the null hypothesis that the NSWD does not have an effect on 1978 salary. We consider the average 1978 income for the group that participated ($5,814) to the average 1978 income for the group that participated ($4,655). The group that participated has an average 1978 income that is $1,159 higher which seems to indicate that the null hypothesis is false. Comparing means suggests that the NSW program did affect 1978 salary.

Next we run a single variable regression with participation in the program (nsw) as the explanatory variable and 1978 earnings (earn_78) as the result. The Excel Regression output is shown below:

Regression Statistics	
Multiple R	0.089704677
R Square	0.008046929
Adjusted R Square	0.006324788
Standard Error	6346.367094
Observations	578

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	188196659.8	188196659.8	4.672631498	0.031058153
Residual	576	23199192169	40276375.29		
Total	577	23387388829			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4654.995685	344.1800798	13.52488409	2.15272E-36
NSWD	1159.42307	536.3659152	2.161627049	0.031058153

We now have further evidence to reject the null hypothesis. Since the p-value of the coefficient (0.31) is less than 0.05, we reject the null hypothesis that the NSWD does not have an effect on 1978 Income.

However, there are other variables which may play a role in the effectiveness of the NSWD program. The inclusion of other explanatory variables might help explain the wage differential and increase the correlation and reduce the standard error of the coefficients. This will enhance the model.

We run next run a regression adding Education and 1975 Earnings as explanatory variables. It may be the case that subjects who have higher education or made more money in the past might be more effectively prepared for the workplace by the NSWD. The following represents the Excel Regression output:

Regression Statistics	
Multiple R	0.171290396
R Square	0.0293404
Adjusted R Square	0.024267266
Standard Error	6288.808916
Observations	578

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	686195338.5	228731779.5	5.783486295	0.000671424
Residual	574	22701193490	39549117.58		
Total	577	23387388829			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	969.7999189	1656.896562	0.585311082	0.558568592
NSWD	1140.196234	531.7540722	2.144217211	0.032434793
Education	316.7002374	158.5263617	1.997776483	0.046212316
1975 Earnings	0.149440585	0.055244785	2.705062306	0.007031951

We see that the P-values for each of the coefficients are smaller than 0.05 and therefore we can interpret NSW participation, Education level, and 1975 Earnings as statistically significant. However, due to the possibility that other variables may be related to the observed wage differential, the model remains subject to endogeneity.

We now consider whether or not married participants are more affected by the NSWD program. It might be the case that subjects with concerns about providing for their spouse might have a greater willingness to learn from the NSWD. We run a regression adding the dummy variable for marital status. The Excel regression follows:

Regression Statistics	
Multiple R	0.172214572
R Square	0.029657859
Adjusted R Square	0.022884092
Standard Error	6293.264766
Observations	578

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	693619877.7	173404969.4	4.378340491	0.001704383
Residual	573	22693768951	39605181.42		
Total	577	23387388829			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	946.2478911	1658.962584	0.570385312	0.568639977
NSWD	1137.959213	532.1559214	2.138394345	0.032907002
Education	315.4666662	158.6642656	1.988265379	0.047257678
1975 Earnings	0.144430315	0.056482034	2.557101895	0.010811275
Marital Status	312.6999459	722.2192451	0.432970941	0.665198806

We see similar results as before concerning the statistical significance of NSWD participation, Education and 1975 Earnings. However, Marital Status has a very large P-Value and is likely not statistically significant.

The original data also contains data concerning age. Perhaps older or younger subjects are more likely to have the NSWD be more effective. The following regression results were produced in Excel:

Regression Statistics	
Multiple R	0.172676357
R Square	0.029817124
Adjusted R Square	0.023044469
Standard Error	6292.748278
Observations	578

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	697344681.7	174336170.4	4.4025752	0.001633897
Residual	573	22690044147	39598680.89		
Total	577	23387388829			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	439.9916407	1935.377589	0.227341498	0.820239297
NSWD	1138.963917	532.0922356	2.140538502	0.032732711
Education	318.8775273	158.6787265	2.009579572	0.044944054
1975 Earnings	0.147252973	0.055432915	2.656417641	0.008117593
Age	20.70297052	39.01650651	0.530620816	0.595887151

Similar to marital status, we see that age does not have a statistical significance to success of the NSWD program measured by 1978 earnings. The P-value for age is very high and should therefore be ignored as a possible predictor for the program's success.

The original data also contains data concerning whether subjects remained in the NSWD until completion. It would be expected that completion would play a large role in the success of the program. We consider a simplified regression to test whether or not completion was a significant predictor. Since the subjects who did not participate in the program could not have dropped out, the regression only considers subjects who participated in the program.

Regression Statistics	
Multiple R	0.093641353
R Square	0.008768703
Adjusted R Square	0.00456857
Standard Error	7218.357677
Observations	238

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	108780025.7	108780025.7	2.087720527	0.149813913
Residual	236	12296706263	52104687.56		
Total	237	12405486289			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	6061.281452	498.1140028	12.16846228	9.16488E-27
Completion	-2098.33292	1452.239394	-1.44489464	0.149813913

The p-value for completion is larger than 0.05 and is therefore statistically insignificant. If we are trying to measure how well the NSWD works for those subjects who complete it, we should include completion in our regression. However, if we are trying to measure if participation in program has a relation to 1978 wages, we should not include completion.

Conclusion

As one might intuitively expect, participation in the NSWD helped people earn more money in 1978. Regressing participation in the NSWD and 1978 earnings shows clear statistical significance. This tells us that the NSWD helped employees find better jobs. Because the data was collected within a randomized experiment, the estimated effect of NSWD is unbiased. Including additional covariates in the model improved its ability to explain observed wage differentials. Specifically, years of education and salary prior to entering the study were shown to be statistically significant predictors of post study employment success. The addition of Age and Marital Status did not improve the accuracy of the model.

The results of our analysis suggest that employee assistance training programs like the NSWD can be effective in increasing the wages of hard-to-employ people.