

Problem Set 4

Question #1: LPM

1.1

```
reg train unem74 unem75 age educ black hisp married
```

| Source | SS | df | MS | Number of obs = | 445 |
|----------|------------|-----|------------|-----------------|--------|
| Model | 2.41922955 | 7 | .345604222 | F(7, 437) = | 1.43 |
| Residual | 105.670658 | 437 | .241809286 | Prob > F = | 0.1915 |
| Total | 108.089888 | 444 | .243445693 | R-squared = | 0.0224 |
| | | | | Adj R-squared = | 0.0067 |
| | | | | Root MSE = | .49174 |

| train | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|---------|-----------|-----------|-------|-------|----------------------|
| unem74 | .02088 | .0772939 | 0.27 | 0.787 | -.1310341 .172794 |
| unem75 | -.0955711 | .0719021 | -1.33 | 0.184 | -.236888 .0457459 |
| age | .0032057 | .0034027 | 0.94 | 0.347 | -.003482 .0098933 |
| educ | .0120131 | .0133419 | 0.90 | 0.368 | -.0142092 .0382354 |
| black | -.0816663 | .0877325 | -0.93 | 0.352 | -.2540963 .0907637 |
| hisp | -.2000168 | .1169708 | -1.71 | 0.088 | -.4299122 .0298785 |
| married | .0372887 | .0644037 | 0.58 | 0.563 | -.0892909 .1638683 |
| _cons | .3380222 | .1894451 | 1.78 | 0.075 | -.0343147 .7103591 |

```
testparm unem74 unem75 age educ black hisp married
```

- (1) unem74 = 0
- (2) unem75 = 0
- (3) age = 0
- (4) educ = 0
- (5) black = 0
- (6) hisp = 0
- (7) married = 0

```
F( 7, 437) = 1.43
Prob > F = 0.1915
```

Since we cannot reject H_0 that $\text{All } \beta_j = 0$, these explanatory variables are not statistically jointly significant at 5% level.

Calculate the percent of correctly predicted:

First, generate dummy variable refer to predicted binary result:

```
. gen Dytrain = 1 if ytrain >=.5
. replace Dytrain = 0 if ytrain <.5
```

```
. tab train Dytrain, cell
```

| =1 if assigned to job training | Dytrain | | Total |
|--------------------------------|--------------|-------------|---------------|
| | 0 | 1 | |
| 0 | 237 53.26 | 23 5.17 | 260 58.43 |
| 1 | 158 35.51 | 27 6.07 | 185 41.57 |
| Total | 395 88.76 | 50 11.24 | 445 100.00 |

```
. display 53.26+6.07
59.33
```

Percent correctly predicted is 59.33%.

1.2 Probit estimation

```
. probit train unem74 unem75 age educ black hisp married
```

```
Iteration 0: log likelihood = -302.1
Iteration 1: log likelihood = -297.01499
Iteration 2: log likelihood = -297.0088
Iteration 3: log likelihood = -297.0088
```

```
Probit regression                               Number of obs = 445
LR chi2(7) = 10.18
Prob > chi2 = 0.1785
Pseudo R2 = 0.0169
Log likelihood = -297.0088
```

| train | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
|---------|-----------|-----------|-------|-------|----------------------|
| unem74 | .0530256 | .1992686 | 0.27 | 0.790 | -.3375337 .4435849 |
| unem75 | -.2477249 | .18505 | -1.34 | 0.181 | -.6104163 .1149665 |
| age | .0083443 | .0087982 | 0.95 | 0.343 | -.0088999 .0255886 |
| educ | .0314431 | .0343238 | 0.92 | 0.360 | -.0358304 .0987165 |
| black | -.2069299 | .2249003 | -0.92 | 0.358 | -.6477264 .2338666 |
| hisp | -.5397772 | .3085029 | -1.75 | 0.080 | -1.144432 .0648773 |
| married | .0966251 | .1655823 | 0.58 | 0.560 | -.2279101 .4211604 |
| _cons | -.4241079 | .4870267 | -0.87 | 0.384 | -1.378663 .5304469 |

```
. est sto ptrain
```

```
. probit train
```

```
Iteration 0: log likelihood = -302.1
Iteration 1: log likelihood = -302.1
```

```
Probit regression                               Number of obs = 445
LR chi2(0) = -0.00
Prob > chi2 = .
Pseudo R2 = -0.0000
Log likelihood = -302.1
```

| train | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
|-------|-----------|-----------|-------|-------|----------------------|
| _cons | -.2128286 | .0599043 | -3.55 | 0.000 | -.3302389 -.0954182 |

```
. est sto p0
```

```
. lrtest ptrain p0
```

```
Likelihood-ratio test                               LR chi2(7) = 10.18
(Assumption: p0 nested in ptrain)                   Prob > chi2 = 0.1785
```

Or by hand: $LR = 2(lur - lr) = 2*(-297.01 - (-302.1)) = 10.18$

We fail to reject H_0 , therefore all variables are not statistically jointly significant.

1.3 Participation in job training can be treated as exogenous for explaining 1978 unemployment status since it is not explained by other observed control variables (unem74,75, age, educ, black, hisp, married) due to its lack of statistical significance. Given these control variables, we can use participation in job training as an exogenous variable together with these control variables in explaining unem78.

1.4 LPM of unem78

```
. reg unem78 train unem74 unem75 age educ black hisp married
```

| Source | SS | df | MS | Number of obs = | 445 |
|----------|------------|-----|------------|-----------------|--------|
| Model | 4.38106947 | 8 | .547633683 | F(8, 436) = | 2.64 |
| Residual | 90.4414024 | 436 | .207434409 | Prob > F = | 0.0078 |
| Total | 94.8224719 | 444 | .213564126 | R-squared = | 0.0462 |
| | | | | Adj R-squared = | 0.0287 |
| | | | | Root MSE = | .45545 |

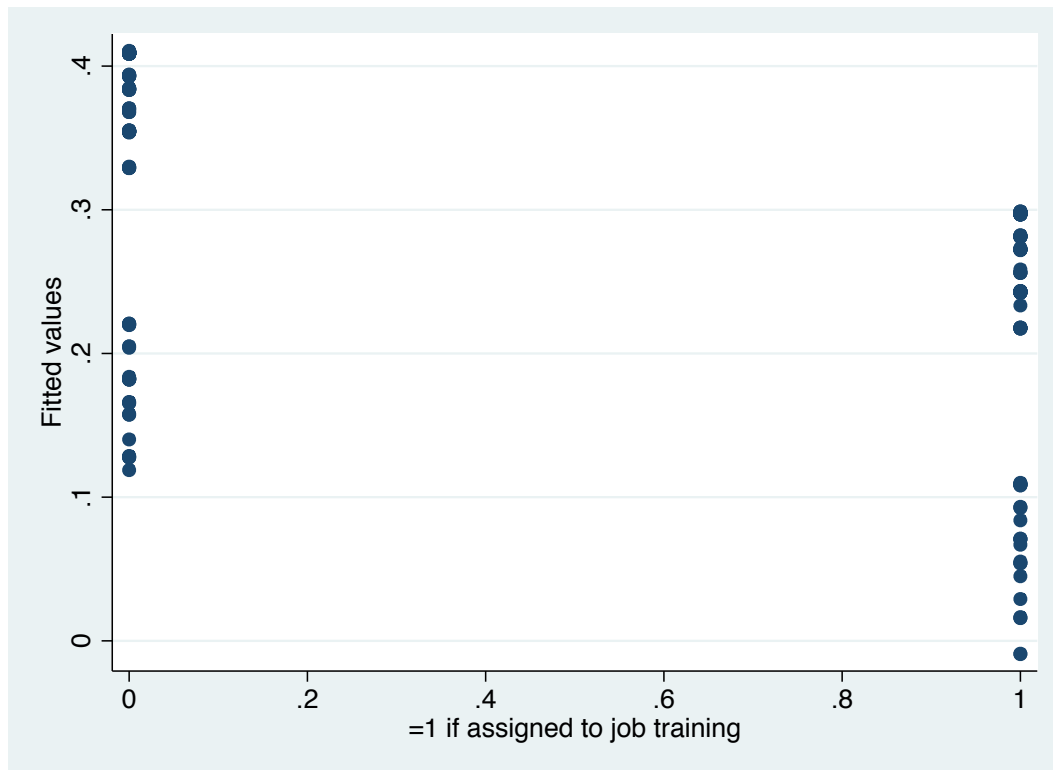
| unem78 | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|---------|-----------|-----------|-------|-------|----------------------|
| train | -.1117028 | .0443061 | -2.52 | 0.012 | -.1987828 -.0246228 |
| unem74 | .0386926 | .0715954 | 0.54 | 0.589 | -.1020226 .1794077 |
| unem75 | .0159613 | .0667301 | 0.24 | 0.811 | -.1151913 .1471139 |
| age | .0000433 | .0031548 | 0.01 | 0.989 | -.0061571 .0062438 |
| educ | .0001442 | .0123687 | 0.01 | 0.991 | -.0241655 .024454 |
| black | .1888328 | .0813382 | 2.32 | 0.021 | .0289692 .3486964 |
| hisp | -.0377011 | .1087 | -0.35 | 0.729 | -.2513423 .1759401 |
| married | -.0254373 | .0596735 | -0.43 | 0.670 | -.1427208 .0918462 |
| _cons | .1631823 | .1761017 | 0.93 | 0.355 | -.1829315 .5092961 |

```
. test train
```

(1) train = 0

F(1, 436) = 6.36
 Prob > F = 0.0121

The coefficient of train is -0.1117, meaning that participating in job training will reduce the probability of being unemployed in 1978 by 0.1117, or 11.17 percentage points, compared to those who did not participate in the job training, holding other variables constant. This estimate is statistically significant at 5% level. See the graph below that a few predicted values are outside [0,1].



1.5 Probit estimation of unem78

It does not make sense to compare the probit coefficient on train with the coefficient obtained from the linear model in (1.4) although we can compare the sign of the coefficient. We can compare the results when we estimate average marginal effects since now we differentiate on predicted probability. From the result below, the average marginal effects of train on unem78 equal to -0.1132, meaning that participating in the job training reduce the probability of being unemployed in 1978 by 11.32 percentage points, ceteris paribus. This is very similar to the result from the linear model.

```
probit unem78 train unem74 unem75 age educ black hisp married
```

```
Iteration 0: log likelihood = -274.73494
Iteration 1: log likelihood = -263.3816
Iteration 2: log likelihood = -263.3128
Iteration 3: log likelihood = -263.31279
```

```
Probit regression                               Number of obs   =       445
                                                LR chi2(8)      =       22.84
                                                Prob > chi2     =       0.0036
Log likelihood = -263.31279                    Pseudo R2      =       0.0416
```

| unem78 | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
|---------|-----------|-----------|-------|-------|----------------------|
| train | -.3365897 | .1316429 | -2.56 | 0.011 | -.5946051 -.0785744 |
| unem74 | .106094 | .2125598 | 0.50 | 0.618 | -.3105155 .5227035 |
| unem75 | .0636124 | .1970995 | 0.32 | 0.747 | -.3226956 .4499204 |
| age | .0006757 | .0091211 | 0.07 | 0.941 | -.0172014 .0185529 |
| educ | -.0018916 | .0367938 | -0.05 | 0.959 | -.0740061 .0702229 |
| black | .6336688 | .2742692 | 2.31 | 0.021 | .0961111 1.171227 |
| hisp | -.1649409 | .3790471 | -0.44 | 0.663 | -.9078596 .5779777 |
| married | -.077768 | .1771557 | -0.44 | 0.661 | -.4249869 .2694509 |
| _cons | -1.010331 | .5380256 | -1.88 | 0.060 | -2.064842 .0441798 |

```
margins, dydx(train)
```

```
Average marginal effects      Number of obs   =      445
Model VCE      : OIM
```

```
Expression      : Pr(unem78), predict()
dy/dx w.r.t.    : train
```

| | dy/dx | Delta-method Std. Err. | z | P> z | [95% Conf. Interval] |
|-------|-----------|---------------------------|-------|-------|------------------------|
| train | -.1131733 | .0433255 | -2.61 | 0.009 | -.1980896 -.0282569 |

Question #2: Tobit model

2.1 There are about 27.92 percent of workers that have pension equal to zero. The range of pension for workers with nonzero pension benefits is between \$7.28 to \$2,880.27. The tobit model is appropriate for modeling pension since we have a nontrivial fraction of the dependent variable with zero value, which is considered as corner solution responses. If we use OLS, we could possibly obtain negative fitted values. Moreover, when the distribution of the dependent variable piles up at zero, we cannot have the dependent variable conditional normal distributed as in a linear model.

```
. gen Dpension = (pension>0 & pension<.)
```

```
. tab Dpension
```

| Dpension | Freq. | Percent | Cum. |
|----------|-------|---------|--------|
| 0 | 172 | 27.92 | 27.92 |
| 1 | 444 | 72.08 | 100.00 |
| Total | 616 | 100.00 | |

```
. sum pension if Dpension == 1
```

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|----------|-----|----------|-----------|------|---------|
| pension | 444 | 905.0439 | 550.3696 | 7.28 | 2880.27 |

2.2 Tobit model

For individual test, male have statistically significant \$308.15 higher expected pension benefits compare to female at 1% significance level. On the other hand, whites does not have statistically significant higher expected pension benefits compare to nonwhites.

For joint test, males and whites have statistically significant higher expected pension benefits than females and nonwhites.

```
. tobit pension exper age tenure educ depends married i.white i.male, ll(0)
```

```
Tobit regression      Number of obs   =      616
LR chi2(8)            =     184.70
Prob > chi2           =     0.0000
Pseudo R2             =     0.0245
Log likelihood = -3672.9635
```

| pension | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] | |
|---------|-----------|-----------|-------|-------|----------------------|-----------|
| exper | 5.203458 | 6.00928 | 0.87 | 0.387 | -6.598007 | 17.00492 |
| age | -4.638944 | 5.710741 | -0.81 | 0.417 | -15.85412 | 6.576228 |
| tenure | 36.02385 | 4.564348 | 7.89 | 0.000 | 27.06005 | 44.98765 |
| educ | 93.21262 | 10.89133 | 8.56 | 0.000 | 71.82343 | 114.6018 |
| depends | 35.28461 | 21.91691 | 1.61 | 0.108 | -7.757432 | 78.32666 |
| married | 53.68858 | 71.73266 | 0.75 | 0.454 | -87.18528 | 194.5624 |
| white | 144.0855 | 102.0753 | 1.41 | 0.159 | -56.37738 | 344.5485 |
| male | 308.1505 | 69.8903 | 4.41 | 0.000 | 170.8948 | 445.4062 |
| _cons | -1252.429 | 219.0692 | -5.72 | 0.000 | -1682.653 | -822.2048 |
| /sigma | 677.7383 | 24.13815 | | | 630.3341 | 725.1426 |

Obs. summary: 172 left-censored observations at pension<=0
444 uncensored observations
0 right-censored observations

. test white male

(1) [model]white = 0
(2) [model]male = 0

F(2, 608) = 11.17
Prob > F = 0.0000

2.3 The difference in expected pension benefits for a white male and a nonwhite female, both of whom are 35 years old, are single with no dependents, have 16 years of education, and have 10 years of experience is calculated below:

(suppose tenure=10)
margins male, at(age=35 married=0 depends=0 tenure=10 educ=16 exper=10 white=1)
predict(ystar(0,.))

Adjusted predictions Number of obs = 616
Model VCE : OIM

Expression : E(pension*|pension>0), predict(ystar(0,.))
at : exper = 10
age = 35
tenure = 10
educ = 16
depends = 0
married = 0
white = 1

| | Margin | Delta-method Std. Err. | z | P> z | [95% Conf. Interval] | |
|------|----------|---------------------------|-------|-------|----------------------|----------|
| male | | | | | | |
| 0 | 696.9033 | 63.7512 | 10.93 | 0.000 | 571.9533 | 821.8534 |
| 1 | 966.6001 | 84.67276 | 11.42 | 0.000 | 800.6445 | 1132.556 |

. margins male, at(age=35 married=0 depends=0 tenure=10 educ=16 exper=10 white=0)
predict(ystar(0,.))

Adjusted predictions Number of obs = 616
Model VCE : OIM

Expression : E(pension*|pension>0), predict(ystar(0,.))

```

at      : exper      =      10
         age        =      35
         tenure     =      10
         educ       =      16
         depends    =       0
         married    =       0
         white      =       0

```

| | Margin | Delta-method Std. Err. | z | P> z | [95% Conf. Interval] | |
|------|----------|---------------------------|------|-------|----------------------|----------|
| male | | | | | | |
| 0 | 582.2655 | 93.03065 | 6.26 | 0.000 | 399.9288 | 764.6023 |
| 1 | 836.9647 | 118.9587 | 7.04 | 0.000 | 603.81 | 1070.119 |

```

display 966.60-582.27
384.33

```

Given other variables as stated above, a white male has \$384.33 more pension benefits than a nonwhite female.

A lot of you guys evaluating tenure at average. Here it is.

```
sum tenure
```

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|----------|-----|----------|-----------|-----|-----|
| tenure | 616 | 7.751623 | 7.775509 | .5 | 25 |

```

margins male, at(age=35 married=0 depends=0 tenure=7.751623 educ=16 exper=10
white=1) predict(ystar(0,.))

```

```

Adjusted predictions      Number of obs   =      616
Model VCE      : OIM

```

```

Expression      : E(pension*|pension>0), predict(ystar(0,.))
at      : exper      =      10
         age        =      35
         tenure     =      7.751623
         educ       =      16
         depends    =       0
         married    =       0
         white      =       1

```

| | Margin | Delta-method Std. Err. | z | P> z | [95% Conf. Interval] | |
|------|----------|---------------------------|-------|-------|----------------------|----------|
| male | | | | | | |
| 0 | 631.3903 | 58.94372 | 10.71 | 0.000 | 515.8627 | 746.9179 |
| 1 | 893.0627 | 80.72283 | 11.06 | 0.000 | 734.8489 | 1051.277 |

```

margins male, at(age=35 married=0 depends=0 tenure=7.751623 educ=16 exper=10
white=0) predict(ystar(0,.))

```

```

Adjusted predictions      Number of obs   =      616
Model VCE      : OIM

```

```

Expression      : E(pension*|pension>0), predict(ystar(0,.))
at      : exper      =      10

```

```

age          =          35
tenure       =       7.751623
educ         =          16
depends       =          0
married      =          0
white        =          0

```

| | Margin | Delta-method Std. Err. | z | P> z | [95% Conf. Interval] | |
|------|----------|---------------------------|------|-------|----------------------|----------|
| male | | | | | | |
| 0 | 521.8628 | 87.10443 | 5.99 | 0.000 | 351.1413 | 692.5844 |
| 1 | 766.6849 | 114.1775 | 6.71 | 0.000 | 542.9011 | 990.4686 |

```

di 893.0627-521.8628
371.1999

```

Given other variables as stated above, a white male has \$371.2 more pension benefits than a nonwhite female.

2.4 Add union to the tobit model

Since we reject H_0 that the coefficient of union is equal to zero at 1% significance level, being in union will increase the *latent* pension benefits by \$439.046, comparing to not being in the union, holding other variables constant.

```

. tobit pension exper age tenure educ depends married white male union, ll(0)

```

```

Tobit regression                               Number of obs   =          616
                                                LR chi2(9)      =          233.52
                                                Prob > chi2     =          0.0000
Log likelihood = -3648.5515                    Pseudo R2       =          0.0310

```

| pension | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] | |
|---------|-----------|-----------|-------|-------|----------------------|-----------|
| exper | 4.393523 | 5.830946 | 0.75 | 0.451 | -7.057754 | 15.8448 |
| age | -1.653532 | 5.555708 | -0.30 | 0.766 | -12.56427 | 9.257211 |
| tenure | 28.77837 | 4.504963 | 6.39 | 0.000 | 19.93116 | 37.62557 |
| educ | 106.8277 | 10.77274 | 9.92 | 0.000 | 85.67134 | 127.9841 |
| depends | 41.46623 | 21.21414 | 1.95 | 0.051 | -.1957922 | 83.12824 |
| married | 19.74554 | 69.50047 | 0.28 | 0.776 | -116.745 | 156.2361 |
| white | 159.2972 | 98.96747 | 1.61 | 0.108 | -35.06298 | 353.6575 |
| male | 257.2457 | 68.02051 | 3.78 | 0.000 | 123.6615 | 390.8298 |
| union | 439.046 | 62.48832 | 7.03 | 0.000 | 316.3265 | 561.7656 |
| _cons | -1571.506 | 218.5445 | -7.19 | 0.000 | -2000.701 | -1142.311 |
| /sigma | 652.8974 | 23.16287 | | | 607.4083 | 698.3865 |

```

Obs. summary:      172 left-censored observations at pension<=0
                   444 uncensored observations
                   0 right-censored observations

```

For a binary independent variable (union), the partial effect should get xb from plugging in union = 1 (with other variables evaluating at mean) and then union = 0. Then calculate $E[y|x, union = 1] - E[y|x, union = 0]$

$$\begin{aligned}
&= \Phi\left(\frac{xb}{\sigma} \middle| union = 1\right) * (xb|union = 1) + \sigma * \phi\left(\frac{xb}{\sigma} \middle| union = 1\right) - \\
&\quad \Phi\left(\frac{xb}{\sigma} \middle| union = 0\right) * (xb|union = 0) + \sigma * \phi\left(\frac{xb}{\sigma} \middle| union = 0\right)
\end{aligned}$$

By hand:

sum exper age tenure educ depends married white male union

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|----------|-----|----------|-----------|-----|-----|
| exper | 616 | 18.61688 | 12.32906 | 0 | 60 |
| age | 616 | 37.65422 | 12.69757 | 16 | 80 |
| tenure | 616 | 7.751623 | 7.775509 | .5 | 25 |
| educ | 616 | 12.51461 | 2.733076 | 6 | 18 |
| depends | 616 | 1.227273 | 1.412017 | 0 | 7 |
| married | 616 | .6866883 | .4642169 | 0 | 1 |
| white | 616 | .9058442 | .2922827 | 0 | 1 |
| male | 616 | .6412338 | .4800282 | 0 | 1 |
| union | 616 | .3181818 | .466149 | 0 | 1 |

```
*union=1, find xb
di 4.3935*18.6169-
1.6535*37.6542+28.7784*7.7516+106.8277*12.5146+41.4662*1.2273+19.7455*0.68
67+159.2972*0.9058+257.2457*0.6412+439.046*1-1571.506
```

820.74476

```
*union=0, find xb
di 4.3935*18.6169-
1.6535*37.6542+28.7784*7.7516+106.8277*12.5146+41.4662*1.2273+19.7455*0.68
67+159.2972*0.9058+257.2457*0.6412+439.046*0-1571.506
```

381.69876

```
di
(normal(820.74476/652.8974)*820.74476+652.8974*normalden(820.74476/652.897
4)) -
(normal(381.69876/652.8974)*381.69876+652.8974*normalden(381.69876/652.897
4))
358.683
```

By STATA, calculating $E[y|x, union = 1]$ and $E[y|x, union = 0]$ while other variables are evaluating at mean.

```
margins, predict(ystar(0,.)) at((mean) _all union=(0 1))
```

```
Adjusted predictions          Number of obs   =          616
Model VCE      : OIM
```

```
Expression      : E(pension*|pension>0), predict(ystar(0,.))
```

```
1._at          : exper          =      18.61688 (mean)
                 age            =      37.65422 (mean)
                 tenure         =       7.751623 (mean)
                 educ           =      12.51461 (mean)
                 depends        =       1.227273 (mean)
                 married        =       .6866883 (mean)
                 white          =       .9058442 (mean)
                 male           =       .6412338 (mean)
                 union          =              0
```

```
2._at          : exper          =      18.61688 (mean)
```

```

age           = 37.65422 (mean)
tenure        = 7.751623 (mean)
educ          = 12.51461 (mean)
depends        = 1.227273 (mean)
married        = .6866883 (mean)
white         = .9058442 (mean)
male          = .6412338 (mean)
union         = 1

```

| | Margin | Delta-method Std. Err. | z | P> z | [95% Conf. Interval] | |
|-----|----------|---------------------------|-------|-------|----------------------|----------|
| _at | | | | | | |
| 1 | 494.6149 | 25.1554 | 19.66 | 0.000 | 445.3112 | 543.9185 |
| 2 | 853.3009 | 44.40107 | 19.22 | 0.000 | 766.2764 | 940.3254 |

```

di 853.3009 - 494.6149
358.686

```

So, the partial effect of being in the union is 358.7. Being in the union provides more pension benefits than not being in the union by \$358.7.