

Applied Microeconometrics, Assignment 3: Matching and Weighting

Sindri Engilbertsson (584872), Ilse van der Voort (584098)

September 2021

Q1: Compute separately for New Jersey and Pennsylvania the average number of employees in both waves, and compute the difference-in-difference estimate. Next repeat this, but only considering the restaurants that responded in both waves of the survey.

We created the new variables:

$$\text{emp_before} = EMPFT + \frac{EMPPT}{2} + NMGRS \quad \text{and} \quad \text{emp_after} = EMPFT + \frac{EMPPT}{2} + NMGRS.$$

We then calculated the following averages for the number of employees in both waves in NJ and PA:

	STATE	emp.before	emp.after
1	NJ	20.44	21.03
2	PA	23.33	21.17

This gives us a difference-in-difference estimator of:

$$NJ_1 - PA_1 - (NJ_0 - PA_0) = 2.7536.$$

So the average number of employees per restaurant in New Jersey increased by 2.7536 employees compared to in Pennsylvania after the increase in minimum wage.

We then repeated the same calculations, but only including those restaurants which responded in both waves of the survey. For this, we included only those restaurants with $STATUS = 1$, as the variable $STATUS$ indicated whether or not a restaurant responded in the second wave. This dropped 11 observations, leaving us with 399 restaurants for which we found the following results:

	STATE	emp.before	emp.after
1	NJ	20.61	21.34
2	PA	23.29	21.52

This time we find a DiD estimate of:

$$NJ_1 - PA_1 - (NJ_0 - PA_0) = 2.5007.$$

We see that this leads to a slightly smaller increase in the average number of employees per restaurant in New Jersey compared to in Pennsylvania, or an increase of 2.5007 employees per restaurant.

Q2: Next we focus on the regression model:

$$E_{1i} - E_{0i} = \alpha + \delta NJ_i + U_i,$$

where $E_{1i} - E_{0i}$ is the change in employment in restaurant i between the two waves, and NJ_i is a dummy variable indicating whether restaurant i is located in New Jersey. Estimate this model and next subsequently add characteristics of the restaurants observed in the first wave. But think carefully which characteristics can be included. How does the latter affect the estimate for the coefficient δ ?

We first run the regression in the simple form provided, i.e. we regress only $STATE$ on the change in employment. The results to that, and the other regressions from this question can be seen in the table below.

In selecting characteristics to include in the larger regression, we first note that we are requested only to include characteristics from the first wave of interviews and thus won't use difference variables.

Next, we know that in order to estimate δ consistently while including individual characteristics, X_i , it is important not to include intermediate outcomes or other bad controls. As we are interested in the relation between an increase in the minimum wage in New Jersey, including the variable PCTAFF, which measures the share of employees affected by new minimum wage, would be incorrect for this reason, as the share affected depends on the size of the wage increase, thus it is an intermediate variable.

Finally, we want the *conditional independence assumption* to hold which states that conditional on X , the outcomes are independent of treatment. So we should include all variables that are relevant for jointly determining treatment and outcomes, which would mean that participation in the treatment program will not depend on outcomes, after controlling for the effect of our controls.

There is no way to test for this assumption, but to find some indication of relation between our treatment variable and the covariates, we regress all covariates on STATE and include significant variables in this sparser regression. All regressions are run using robust standard errors. The results are as follows:

Table 1: Three OLS regressions with robust standard errors.

	<i>Dependent variable:</i>		
	emp_diff		
	(1)	(2)	(3)
STATE	2.722** (1.335)	−0.012 (2.429)	0.748 (2.122)
CENTRALJ		−1.940 (1.818)	−1.505 (1.549)
NORTHJ		−1.341 (1.421)	−1.177 (1.233)
SHORE		−2.284 (1.835)	−2.335 (1.749)
PA1		−4.537 (3.071)	−3.796 (2.598)
CHAIN		−0.821 (0.563)	−0.535 (0.488)
CO-OWNED		0.502 (1.194)	
NCALLS		0.213 (0.327)	
WAGE_ST		−2.438 (1.668)	−2.722* (1.464)
INCTIME		0.075 (0.051)	
FIRSTINC		−0.167 (6.165)	
BONUS		−0.182 (1.237)	
MEALS		0.161 (0.948)	1.136 (0.877)
OPEN		1.218 (0.784)	1.073 (0.674)
HRSOPEN		0.413 (0.574)	0.353 (0.517)
PSODA		0.823 (9.194)	−1.563 (8.307)
PFRY		8.990 (7.462)	9.382 (6.829)
PENTREE		−0.684 (1.107)	
NREGS		0.384 (0.534)	0.161 (0.366)
NREGS11		−0.577 (0.655)	
Constant	−2.064* (1.246)	−11.485 (18.352)	−9.374 (16.145)
Observations	378	292	340
R ²	0.015	0.065	0.064
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01			

We see that the base regression gives results similar to the difference-in-difference estimator, while the regressions including more covariates have a smaller and not significant coefficient for the STATE variable. The assumptions necessary for these OLS regressions to be valid are, however, unlikely to hold, as we will see in the following questions.

Q3: Provide a balancing table, i.e. show the sample mean of characteristics observed in the first survey separately for the restaurants in New Jersey and Pennsylvania. What is your opinion about the balancing table?

	n_0	mean_0	sd_0	n_1	mean_1	sd_1	Diff
CHAIN	79	2.15	1.19	331	2.11	1.09	-0.043
CO_OWNED	79	0.35	0.48	331	0.34	0.47	-0.013
Employees_before	77	23.33	11.86	321	20.44	9.11	-2.892**
Employees_after	77	21.17	8.28	319	21.03	9.29	-0.138
DEmployees	75	-2.28	10.85	309	0.47	8.45	2.750**
NCALLS	79	0.78	0.97	331	1.23	1.46	0.442**
EMPFT	78	10.21	10.78	326	7.72	7.96	-2.481**
EMPPT	78	19.49	9.58	328	18.68	10.21	-0.812
NMGRS	78	3.54	1.10	326	3.39	1.00	-0.146
WAGE_ST	76	4.63	0.35	314	4.61	0.35	-0.018
INCTIME	74	18.87	13.10	305	18.17	10.71	-0.698
FIRSTINC	69	0.21	0.10	298	0.23	0.11	0.022
BONUS	79	0.29	0.46	331	0.24	0.43	-0.055
PCTAFF	74	45.91	36.09	292	49.62	34.89	3.715
MEALS	79	2.03	0.39	331	1.88	0.56	-0.146**
OPEN	79	7.82	2.16	331	8.10	2.16	0.279
HRSOPEN	79	14.53	2.95	331	14.42	2.78	-0.107
PSODA	77	0.97	0.07	325	1.06	0.08	0.087***
PFRY	77	0.84	0.09	316	0.94	0.10	0.100***
PENTREE	77	1.22	0.62	321	1.35	0.65	0.133
NREGS	78	3.33	1.10	326	3.66	1.29	0.323**
NREGS11	77	2.79	0.75	321	2.70	0.92	-0.094

To use DiD we need a common trend. However, we do not have sufficient information on previous periods to check for this common trend. We thus look at the balancing table to compare the groups before the minimum wage increase. The first thing we note is that there is a big difference in the number of observations in both states. This gives extra reason to check for balanced characteristics. We then note that, while most of the characteristic variables are balanced between the two states, there is a significant difference in the number of employees in the first wave. We also notice unbalancedness in other characteristics, such as the price of soda. This indicates that the groups were not properly balanced. We need to take this into account when computing the propensity scores.

We are not surprised by the fact that not all characteristics are balanced between the treatment and the control group. This is a natural result of the fact that the groups are not randomized. Even though these are neighbouring states, there are obvious reasons for why there will be systematic differences between them, for instance, due to different tax laws.

Q4: Check for the different characteristics if there is a common support for restaurants in New Jersey and Pennsylvania. And estimate a propensity score for being a restaurant in New Jersey.

There are few variables where all values are existent in both states, namely BONUS, CHAIN, and COOWNED. For all other variables there are values that only exist in one state and therefore in those cases perfectly predict treatment participation. This can be seen in the attached file where tables with frequency counts are shown for all variables that could possibly be used in calculating the propensity scores. Most of these variables have near continuous values, and so this result is not surprising. This indicates that inexact matching methods are suitable.

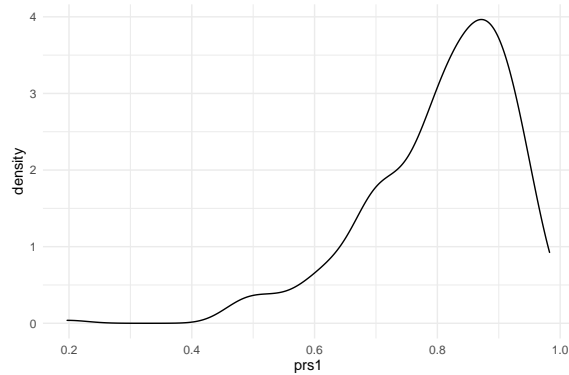
We exclude the variable PCTAFF because it is an intermediate variable. We also exclude the variables that make up the number of employees. We do not include location dummies as they are perfectly indicative of treatment. We then estimate the propensity scores by a logit model. We estimate two versions of our logit model. One model with all the balanced variables from the table, and one model with the balanced variables that are also significant in regressing STATE on all characteristics. The models give the following results:

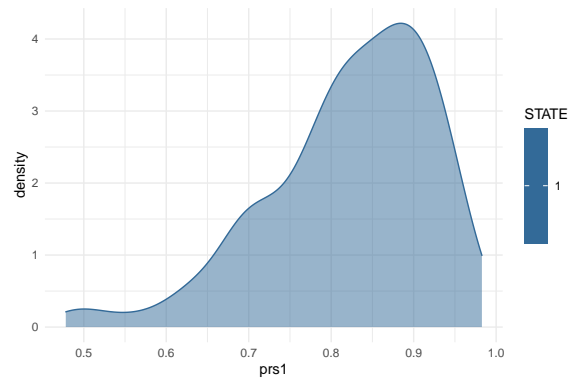
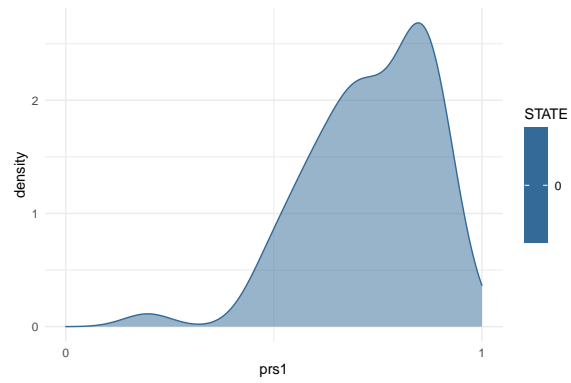
Table 2: Two logit models.

	<i>Dependent variable:</i>	
	STATE	
	(1)	(2)
WAGE_ST	−0.439 (0.460)	−0.120 (0.397)
INCTIME	−0.017 (0.015)	
FIRSTINC	2.907* (1.590)	
BONUS	−0.531 (0.342)	
OPEN	0.643*** (0.214)	0.425** (0.168)
HRSOPEN	0.537*** (0.160)	0.278** (0.135)
PENTREE	0.591* (0.303)	
NREGS11	−0.064 (0.156)	
CHAIN	−0.085 (0.153)	−0.062 (0.130)
CO_OWNED	−0.246 (0.307)	
Constant	−9.853** (3.995)	−5.316 (3.538)
Observations	312	379
Log Likelihood	−141.094	−185.394
Akaike Inf. Crit.	304.189	380.788

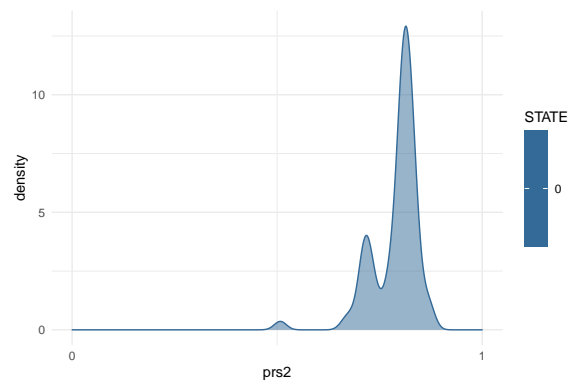
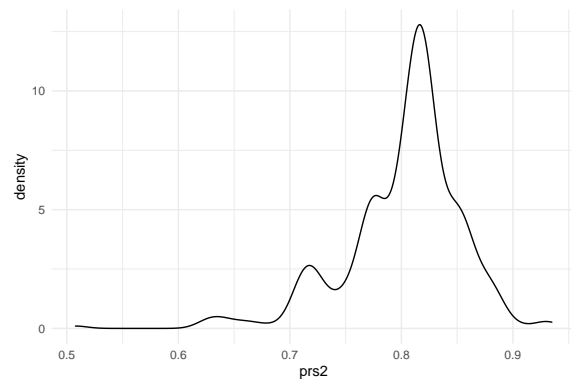
Note: *p<0.1; **p<0.05; ***p<0.01

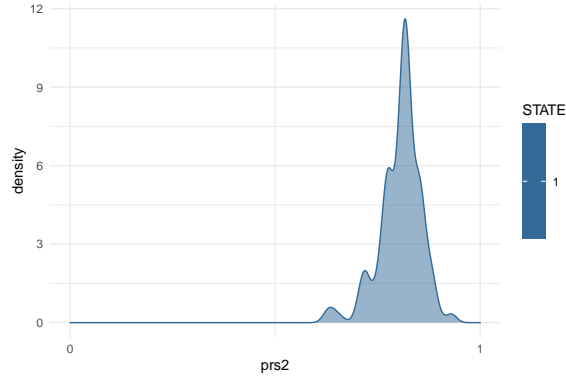
We use the models to predict the propensity scores and check how much they overlap. The following graphs show the distribution of the propensity scores for the first logit model for both states combined and for each state individually.





The following graphs show the distribution of the propensity scores for the first logit model for both states combined and for each state individually.





The plots show that, for both model versions, there is significant overlap between the propensity scores for the two states. There is no clear distinction in the scores for restaurants in NJ versus those in PA. This indicates weakness in the propensity scores.

Q5: Use propensity score matching to estimate the average treatment effect on the treated for the employment before and after the minimum wage increase in New Jersey, so on E_{0i} and E_{1i} separately.

We follow what we have done above and calculate the results for two separate models, a larger model, including all of the balanced variables according to the balancing table in Q3, as well as a parser model, including only those variables that are both balanced and significant w.r.t. the state variable.

Using the MatchIt function in R, we create new "matched" data frames based on the propensity scores for our two models, sacrificing a lot of our observations in the process. We then use these data as inputs for the original base regression of STATE on employment before and after the change, using robust standard errors.

We find that the estimates are significant in the first wave at a 1 percent level and insignificant for the second wave at a 10 percent level. The mean difference between the E_{0i} and E_{1i} for the larger model is equal to 3.322 and the difference for the sparser model is 3.363. These numbers are slightly higher than for our DiD estimates.

The results to these regressions can be seen in Table 3 below.

Q6: Now use propensity score matching to estimate the average treatment effect on the treated on the change in employment in the restaurants, so $E_{1i} - E_{0i}$.

Here we do the same as in Q5, adding a regression on the difference between employment before and after the minimum wage hike. The results to all the regressions are in the following table:

Table 3: Propensity Score Matching Regressions

	<i>Dependent variable:</i>					
	emp_before		emp_after		emp.diff	
	Larger Model	Parser Model	Larger Model	Parser Model	Larger Model	Parser Model
	(1)	(2)	(3)	(4)	(5)	(6)
STATE	-4.429** (1.834)	-4.504*** (1.705)	-1.107 (1.525)	-0.364 (1.317)	3.368* (1.776)	3.937** (1.629)
Constant	23.238*** (1.546)	23.497*** (1.383)	21.415*** (1.023)	21.695*** (0.928)	-1.935 (1.435)	-1.915 (1.286)
Observations	120	144	119	147	117	141
R ²	0.047	0.046	0.004	0.001	0.030	0.040

Note:

*p<0.1; **p<0.05; ***p<0.01

We see that the results still indicate an increase in employment in NJ compared to PA following the hike in minimum wages in NJ. The results are even larger than those from our difference-in-difference estimate. We will note that we are left with less than 150 observations after applying the matching.

Q7: Now check the sensitivity of the propensity score matching estimate by also computing the weighting estimators for the average treatment effect on the treated.

We use the the following formula to calculate the ATET:

$$ATET = \frac{\sum_{i=1}^n \hat{p}(X_i) \left(\frac{D_i Y_i}{\hat{p}(X_i)} - \frac{(1-D_i) Y_i}{1-\hat{p}(X_i)} \right)}{\sum_{i=1}^n \hat{p}(X_i)}$$

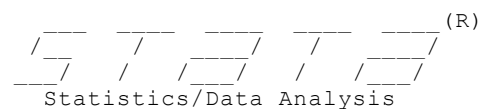
D_i indicates STATE, Y_i indicates the number of employees either in the first wave, second wave, or as a difference. We calculate E_0 and E_1 separately for both sets of propensity scores and get the following:

	E_0	E_1	$E_1 - E_0$
1	-5.144186	-2.803461	2.340724
2	-3.373046	-0.7689948	2.604051

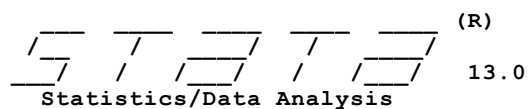
Note that to get the values in the table above we used the number of employees to get Y_i . We also calculate $E_{1i} - E_{0i}$ directly where we use the difference in the number of employees for Y_i . We then get the following:

	$E_1 - E_0$
1	2.757367
2	2.43858

All different ways to estimate the ATET give similar results, which are also similar to de DiD estimator with all observations (2.7536) and with only those who responded in both waves (2.5007). The results thus seem robust, although they might differ greatly with other propensity scores.



User: Stata Logfile
Project: Stata Logfile



MP - Parallel Edition

Copyright 1985-2013 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

3-user 8-core Stata network perpetual license:
Serial number: 501306208483
Licensed to: IDRE-UCLA
IDRE-UCLA

Notes:

1. (/v# option or -set maxvar-) 5000 maximum variables

```
1 . do "C:\Users\Ilse\AppData\Local\Temp\STD01000000.tmp"
2 . use "C:\Users\Ilse\Dropbox\Alles\Alles Dropbox\TI\2.1\Applied Microeconometrics\Assignment_3\
3 .
   end of do-file
4 . do "C:\Users\Ilse\AppData\Local\Temp\STD01000000.tmp"
5 . gen Employees1 = EMPFT + EMPPT/2 + NMGRS
   (12 missing values generated)
6 . gen Employees2 = EMPFT2 + EMPPT2/2 + NMGRS2
   (14 missing values generated)
7 . gen Employees_before = Employees1
   (12 missing values generated)
8 . gen Employees_after = Employees2
   (14 missing values generated)
9 .
   end of do-file
10 . do "C:\Users\Ilse\AppData\Local\Temp\STD01000000.tmp"
11 . gen DEmployees = Employees2-Employees1
   (26 missing values generated)
12 .
   end of do-file
13 . do "C:\Users\Ilse\AppData\Local\Temp\STD01000000.tmp"
14 . tab NCALLS STATE
```

NCALLS	STATE		Total
	0	1	
0	46	162	208
1	5	32	37
2	27	68	95
3	1	52	53
4	0	11	11
5	0	2	2
6	0	2	2
8	0	2	2
Total	79	331	410

15 . tab WAGE_ST STATE

WAGE_ST	STATE		Total
	0	1	
4.25	26	101	127
4.32	0	1	1
4.35	1	5	6
4.37	0	5	5
4.39	0	1	1
4.4	0	1	1
4.45	0	2	2
4.5	13	50	63
4.55	0	2	2
4.6	0	2	2
4.62	0	15	15
4.65	0	2	2
4.67	1	1	2
4.75	10	38	48
4.8	0	4	4
4.85	0	2	2
4.87	1	7	8
4.95	0	2	2
5	20	45	65
5.05	0	4	4
5.06	0	1	1
5.1	0	1	1
5.12	0	3	3
5.15	0	1	1
5.25	2	6	8
5.3	0	1	1
5.37	0	1	1
5.42	0	1	1
5.5	2	6	8
5.56	0	1	1
5.62	0	1	1
5.75	0	1	1
Total	76	314	390

16 . tab INCTIME STATE

INCTIME	STATE		Total
	0	1	
2	1	2	3
2.5	0	1	1
3	0	1	1
3.5	1	0	1
4	11	31	42
6	1	6	7
6.5	0	1	1
7	0	1	1
8	1	9	10
9	1	11	12
10	0	3	3
11	0	2	2
12	0	1	1
13	24	93	117
14	0	1	1
15	0	2	2
16	0	3	3
17	2	3	5
18	1	5	6
19	4	4	8
20	0	1	1
21	0	1	1
26	19	105	124
30	0	1	1
39	2	6	8
52	6	11	17

Total	74	305	379
-------	-----------	------------	------------

17 . tab FIRSTINC STATE

FIRSTINC	STATE		Total
	0	1	
.05	0	3	3
.07	5	0	5
.1	5	18	23
.12	6	27	33
.13	0	1	1
.15	13	46	59
.17	5	19	24
.18	0	1	1
.2	4	26	30
.21	0	1	1
.22	1	5	6
.25	18	91	109
.27	0	13	13
.3	2	5	7
.32	0	5	5
.33	0	1	1
.35	0	1	1
.36	0	1	1
.37	9	16	25
.4	0	1	1
.42	0	1	1
.5	1	7	8
.52	0	3	3
.57	0	1	1
.62	0	1	1
.66	0	1	1
.67	0	1	1
.75	0	2	2
Total	69	298	367

18 . tab BONUS STATE

BONUS	STATE		Total
	0	1	
0	56	253	309
1	23	78	101
Total	79	331	410

19 . tab MEALS STATE

MEALS	STATE		Total
	0	1	
0	0	1	1
1	5	71	76
2	67	226	293
3	7	33	40
Total	79	331	410

20 . tab OPEN STATE

OPEN	STATE		Total
	0	1	
0	1	4	5
5	0	3	3
6	23	61	84
6.5	13	12	25
7	10	113	123
7.5	1	2	3
8	0	8	8
8.5	1	3	4
9	3	7	10
10	12	11	23
10.5	6	51	57
11	9	55	64
11.5	0	1	1
Total	79	331	410

21 . tab HRSOPEN STATE

HRSOPEN	STATE		Total
	0	1	
7	0	1	1
8	1	0	1
9.5	0	1	1
10	8	25	33
10.5	1	11	12
11	0	20	20
11.5	6	26	32
12	9	13	22
12.5	4	21	25
13	3	4	7
13.5	0	6	6
14	2	6	8
14.5	0	4	4
15	3	28	31
15.5	1	4	5
16	15	86	101
16.5	12	8	20
17	2	46	48
17.5	1	3	4
18	9	8	17
18.5	0	1	1
19	1	5	6
24	1	4	5
Total	79	331	410

22 . tab PSODA STATE

PSODA	STATE		Total
	0	1	
.73	1	0	1
.84	1	0	1
.85	0	3	3
.87	5	0	5
.89	4	1	5
.9	1	1	2
.91	3	14	17
.92	1	1	2
.93	0	1	1
.94	7	0	7
.95	10	30	40
.96	0	4	4
.97	9	0	9

.98	5	1	6
.99	0	1	1
1	1	3	4
1.01	6	3	9
1.02	1	34	35
1.03	3	1	4
1.04	1	4	5
1.05	15	5	20
1.06	0	116	116
1.07	0	4	4
1.09	0	9	9
1.1	0	7	7
1.11	2	1	3
1.12	0	33	33
1.13	1	2	3
1.15	0	1	1
1.16	0	1	1
1.17	0	15	15
1.18	0	6	6
1.19	0	3	3
1.2	0	5	5
1.22	0	2	2
1.23	0	4	4
1.26	0	1	1
1.27	0	6	6
1.28	0	1	1
1.49	0	1	1
Total	77	325	402

23 . tab PFRY STATE

PFRY	STATE		Total
	0	1	
.67	0	1	1
.69	1	1	2
.7	0	1	1
.71	0	1	1
.72	1	0	1
.73	6	0	6
.74	6	10	16
.75	3	1	4
.77	4	0	4
.78	1	0	1
.79	5	5	10
.8	2	10	12
.81	0	2	2
.82	4	0	4
.83	2	1	3
.84	12	3	15
.85	0	36	36
.86	3	9	12
.87	1	2	3
.88	1	6	7
.89	4	8	12
.9	5	1	6
.91	5	27	32
.92	0	5	5
.93	0	6	6
.94	2	4	6
.95	1	68	69
.96	0	1	1
.97	0	1	1
.98	1	9	10
.99	0	1	1
1	0	8	8
1.01	4	8	12
1.02	0	32	32
1.03	1	1	2
1.04	0	1	1

1.05	2	1	3
1.06	0	20	20
1.07	0	10	10
1.09	0	1	1
1.12	0	4	4
1.17	0	4	4
1.18	0	1	1
1.19	0	1	1
1.27	0	4	4
Total	77	316	393

24 . tab PENTREE STATE

PENTREE	STATE		Total
	0	1	
.49	0	1	1
.52	3	1	4
.63	1	2	3
.73	1	0	1
.74	0	2	2
.79	2	2	4
.8	3	2	5
.81	1	0	1
.82	0	2	2
.84	5	1	6
.85	2	10	12
.86	4	0	4
.87	0	1	1
.88	0	2	2
.89	1	4	5
.9	2	1	3
.91	6	12	18
.92	0	1	1
.93	0	5	5
.94	11	3	14
.95	4	61	65
.96	2	4	6
.98	0	7	7
.99	1	2	3
1	0	2	2
1.01	3	6	9
1.02	0	15	15
1.03	0	3	3
1.04	0	8	8
1.05	4	1	5
1.06	0	57	57
1.07	0	2	2
1.08	0	1	1
1.09	0	1	1
1.1	0	5	5
1.11	1	3	4
1.12	0	4	4
1.16	2	0	2
1.17	0	4	4
1.38	0	1	1
1.5	0	1	1
1.65	1	0	1
1.75	1	0	1
1.79	1	0	1
1.8	1	0	1
1.94	0	1	1
2.01	2	0	2
2.02	0	1	1
2.03	0	1	1
2.13	0	6	6
2.14	0	1	1
2.19	0	3	3
2.24	0	1	1
2.28	0	1	1

2.3	0	3	3
2.32	3	0	3
2.33	3	0	3
2.34	0	30	30
2.35	2	1	3
2.43	1	0	1
2.45	0	6	6
2.49	0	1	1
2.53	2	1	3
2.56	0	8	8
2.57	0	1	1
2.61	0	1	1
2.66	0	7	7
2.74	0	1	1
2.77	0	4	4
3.09	0	1	1
3.5	1	0	1
3.95	0	1	1
Total	77	321	398

25 . tab NREGS STATE

NREGS	STATE		Total
	0	1	
1	0	1	1
2	19	60	79
3	29	104	133
4	19	84	103
5	7	51	58
6	4	18	22
7	0	4	4
8	0	4	4
Total	78	326	404

26 . tab NREGS11 STATE

NREGS11	STATE		Total
	0	1	
1	3	20	23
2	21	123	144
3	43	126	169
4	9	39	48
5	1	12	13
6	0	1	1
Total	77	321	398

27 . tab CHAIN STATE

CHAIN	STATE		Total
	0	1	
1	35	136	171
2	12	68	80
3	17	82	99
4	15	45	60
Total	79	331	410

28 . tab CO_OWNED STATE

CO_OWNED	STATE		Total
	0	1	
0	51	218	269
1	28	113	141
Total	79	331	410

29 .
end of do-file

30 .
31 .
32 .
33 .
34 .
35 .
36 .
37 .
38 .
39 .
40 .
41 .
42 .
43 .
44 .
45 .
46 .
47 .
48 .
49 .
50 .
51 .
52 .
53 .
54 .
55 .
56 .
57 .
58 .
59 .
60 .
61 .