

ELV-781 Project Report:

Submitted By:

Udit Raj Singh 2016EE10469

Nishant Verma 2016EE10471

Abbreviations:

agem1	age
agefstm	Age at first birth
educm	Years of education
whitem	white
blackm	black
hispn	Hispanic
othracem	Other Race

workedm	If worked for pay
weeksm1	weeks worked
hourswm	hours worked
incomem	annual labour income
famincl	log of family income

Means and Standard deviations:

Xi Mean	All women	wife	husband
Age	29.868	29.97461	32.65742
age at first birth	20.49905	20.74286	23.68556
Years of education	12.1632	12.26449	12.82855
White	0.822795	0.866712	0.866415
Black	0.117915	0.075778	0.078124
Hispanic	0.030149	0.027383	0.028169
Other race	0.029142	0.030128	0.027292

Xi Std	All women	wife	husband
age	3.62831	3.574342	5.064661
age at first birth	3.069067	3.076657	4.512551
Years of education	2.412824	2.4147	3.087226
White	0.381843	0.339887	0.340207
Black	0.322508	0.264643	0.268368
Hispanic	0.170997	0.163196	0.165454
Other racem	0.168204	0.170939	0.162933

Yi mean	All women	wife	husband
Worked for pay	0.557496	0.540688	0.974274
Weeks worked	20.24203	19.32397	47.47637

Hours worked	18.49821	17.54797	43.17618
Annual labour Income	37192.18	37192.18	6506.149
Log family income	10.31362	10.52635	10.31752

Yi Std	All women	wife	husband
Worked for pay	0.496684	0.498342	0.158316
Weeks worked	22.09244	21.83174	11.09071
Hours worked	18.89402	18.63055	12.4738
Annual labour Income	24360.84	24360.84	10345.61
Log family income	1.348681	1.046161	1.261005

Wi (More than 2 kids)	All women	wife
mean	0.2708637318055	0.2816608397796
Standard Deviation	0.4444061021024	0.4498091554687

OLS

OLS	All women	wife	husband
Worked for pay	-0.16813554	-0.16813554	-0.00538576431
Weeks worked	-9.71246319	-8.73406347	-0.523228315
Hours worked	-7.55095385	-6.78612904	0.386591204
Annual labour Income	4181.51824	1180.97577	659.135724
Log family income	-0.0601087233	-0.0965953603	-0.116319706

Honest estimation:

Normalised MSE	All women	Married Women	Husbands
Worked for pay	0.38547055103220407	0.4114240813643962	0.026642579912253612
Weeks worked	0.4750198186260566	0.5038574416539014	0.0500624673279833
Hours worked	0.45263548477360244	0.48017817928095047	0.07586019148016028
Annual labour Income	0.37726226588236605	0.2777415084530052	0.6467853138036036
Log family income	0.01481573410354812	0.008831986434439016	0.014136053755134338

Decision Trees

Plain decision trees on treatment 0 and treatment 1 data

Treatment = 1

All Women	Worked for pay	
All Women	Weeks worked	
All Women	Hours worked	
All Women	Annual labour Income	

All Women	Log family income	
-----------	-------------------	--

Treatment =0

All Women	Worked for pay	
All Women	Weeks worked	
All Women	Hours worked	
All Women	Annual labour Income	

All Women	Log family income	<pre> graph TD Root["educm <= 11.5 mse = 1.678 samples = 676104 value = 10.363"] Root -- True --> Node1["blackm <= 0.5 mse = 2.648 samples = 110914 value = 9.882"] Root -- False --> Node2["agem1 <= 27.5 mse = 1.43 samples = 565490 value = 10.481"] Node1 -- True --> Node3["agem1 <= 27.5 mse = 2.336 samples = 95251 value = 9.989"] Node1 -- False --> Node4["agem1 <= 28.5 mse = 4.074 samples = 15363 value = 9.222"] Node2 -- True --> Node5["blackm <= 0.5 mse = 1.608 samples = 167131 value = 10.19"] Node2 -- False --> Node6["educm <= 15.5 mse = 1.305 samples = 398359 value = 10.603"] Node3 -- True --> Node7["mse = 2.429 samples = 45571 value = 9.842"] Node3 -- False --> Node8["mse = 2.213 samples = 49680 value = 10.123"] Node4 -- True --> Node9["mse = 4.416 samples = 7307 value = 9.028"] Node4 -- False --> Node10["mse = 3.699 samples = 8056 value = 9.398"] Node5 -- True --> Node11["mse = 1.352 samples = 141815 value = 10.271"] Node5 -- False --> Node12["mse = 2.797 samples = 25316 value = 9.738"] Node6 -- True --> Node13["mse = 1.377 samples = 303138 value = 10.522"] Node6 -- False --> Node14["mse = 0.99 samples = 95221 value = 10.861"] </pre>
-----------	-------------------	--

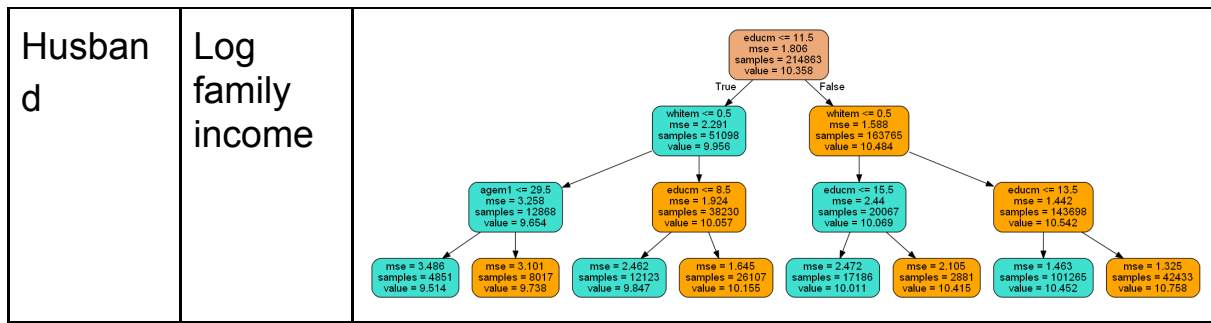
Treatment = 1

Married Women	Worked for pay	<pre> graph TD Root["blackm <= 0.5 mse = 0.25 samples = 214863 value = 0.507"] Root -- True --> Node1["agem1 <= 32.5 mse = 0.25 samples = 196655 value = 0.493"] Root -- False --> Node2["educm <= 11.5 mse = 0.224 samples = 5574 value = 0.662"] Node1 -- True --> Node3["agefstm <= 20.5 mse = 0.245 samples = 82635 value = 0.432"] Node1 -- False --> Node4["agefstm <= 22.5 mse = 0.249 samples = 114020 value = 0.536"] Node2 -- True --> Node5["agem1 <= 28.5 mse = 0.246 samples = 5574 value = 0.56"] Node2 -- False --> Node6["educm <= 12.5 mse = 0.207 samples = 12634 value = 0.707"] Node3 -- True --> Node7["mse = 0.249 samples = 56194 value = 0.491"] Node3 -- False --> Node8["mse = 0.233 samples = 23451 value = 0.37"] Node4 -- True --> Node9["mse = 0.244 samples = 63060 value = 0.58"] Node4 -- False --> Node10["mse = 0.25 samples = 50960 value = 0.483"] Node5 -- True --> Node11["mse = 0.25 samples = 2005 value = 0.499"] Node5 -- False --> Node12["mse = 0.241 samples = 3569 value = 0.595"] Node6 -- True --> Node13["mse = 0.219 samples = 5221 value = 0.675"] Node6 -- False --> Node14["mse = 0.179 samples = 4413 value = 0.768"] </pre>
Married Women	Weeks worked	<pre> graph TD Root["agem1 <= 32.5 mse = 470.424 samples = 214863 value = 18.22"] Root -- True --> Node1["blackm <= 0.5 mse = 412.18 samples = 93073 value = 15.03"] Root -- False --> Node2["agefstm <= 22.5 mse = 501.221 samples = 121790 value = 20.657"] Node1 -- True --> Node3["agem1 <= 27.5 mse = 390.693 samples = 82635 value = 13.921"] Node1 -- False --> Node4["agem1 <= 28.5 mse = 495.458 samples = 10438 value = 23.811"] Node2 -- True --> Node5["blackm <= 0.5 mse = 513.346 samples = 68301 value = 22.744"] Node2 -- False --> Node6["agem1 <= 37.5 mse = 473.075 samples = 53489 value = 17.992"] Node3 -- True --> Node7["mse = 310.733 samples = 24535 value = 10.995"] Node3 -- False --> Node8["mse = 419.316 samples = 58100 value = 15.157"] Node4 -- True --> Node9["mse = 435.774 samples = 3034 value = 18.216"] Node4 -- False --> Node10["mse = 501.834 samples = 7404 value = 26.103"] Node5 -- True --> Node11["mse = 508.534 samples = 63060 value = 22.079"] Node5 -- False --> Node12["mse = 501.827 samples = 5241 value = 30.75"] Node6 -- True --> Node13["mse = 424.662 samples = 23064 value = 15.028"] Node6 -- False --> Node14["mse = 498.068 samples = 30425 value = 20.239"] </pre>
Married Women	Hours worked	<pre> graph TD Root["blackm <= 0.5 mse = 344.247 samples = 214863 value = 16.273"] Root -- True --> Node1["agefstm <= 20.5 mse = 335.927 samples = 196655 value = 15.559"] Root -- False --> Node2["educm <= 11.5 mse = 369.102 samples = 18208 value = 23.988"] Node1 -- True --> Node3["agem1 <= 29.5 mse = 358.638 samples = 91818 value = 17.2"] Node1 -- False --> Node4["agem1 <= 34.5 mse = 311.614 samples = 104837 value = 14.122"] Node2 -- True --> Node5["agem1 <= 28.5 mse = 402.392 samples = 5574 value = 20.3"] Node2 -- False --> Node6["educm <= 12.5 mse = 345.777 samples = 12634 value = 25.612"] Node3 -- True --> Node7["mse = 337.639 samples = 33283 value = 13.956"] Node3 -- False --> Node8["mse = 361.196 samples = 58535 value = 19.044"] Node4 -- True --> Node9["mse = 277.025 samples = 41347 value = 11.323"] Node4 -- False --> Node10["mse = 325.717 samples = 63490 value = 15.944"] Node5 -- True --> Node11["mse = 403.363 samples = 2005 value = 17.959"] Node5 -- False --> Node12["mse = 397.038 samples = 3569 value = 21.616"] Node6 -- True --> Node13["mse = 355.788 samples = 8221 value = 24.423"] Node6 -- False --> Node14["mse = 319.584 samples = 4413 value = 27.828"] </pre>
Married Women	Annual labour Income	<pre> graph TD Root["educm <= 13.5 mse = 83538047.392 samples = 214863 value = 39806.362"] Root -- True --> Node1["educm <= 11.5 mse = 58683102.781 samples = 18678 value = 35496.421"] Root -- False --> Node2["agem1 <= 33.5 mse = 81590730.977 samples = 17450 value = 44842.06"] Node1 -- True --> Node3["whitem <= 0.5 mse = 39420440.479 samples = 51098 value = 27121.867"] Node1 -- False --> Node4["agem1 <= 31.5 mse = 62788519.198 samples = 39195.297"] Node2 -- True --> Node5["whitem <= 0.5 mse = 160304017.893 samples = 30825 value = 60409.578"] Node2 -- False --> Node6["educm <= 13.5 mse = 1286533727.165 samples = 3215 value = 45541.062"] Node3 -- True --> Node7["mse = 308920030.34 samples = 12688 value = 21904.834"] Node3 -- False --> Node8["mse = 411338053.101 samples = 35232 value = 28877.89"] Node4 -- True --> Node9["mse = 426923508.303 samples = 41103 value = 33633.443"] Node4 -- False --> Node10["mse = 709950254.173 samples = 1452 value = 42280.293"] Node5 -- True --> Node11["mse = 586788425.275 samples = 4526 value = 35853.278"] Node5 -- False --> Node12["mse = 1005429030.881 samples = 12524 value = 47972.531"] Node6 -- True --> Node13["mse = 1286533727.165 samples = 3215 value = 45541.062"] Node6 -- False --> Node14["mse = 1613223821.553 samples = 27405 value = 62155.975"] </pre>

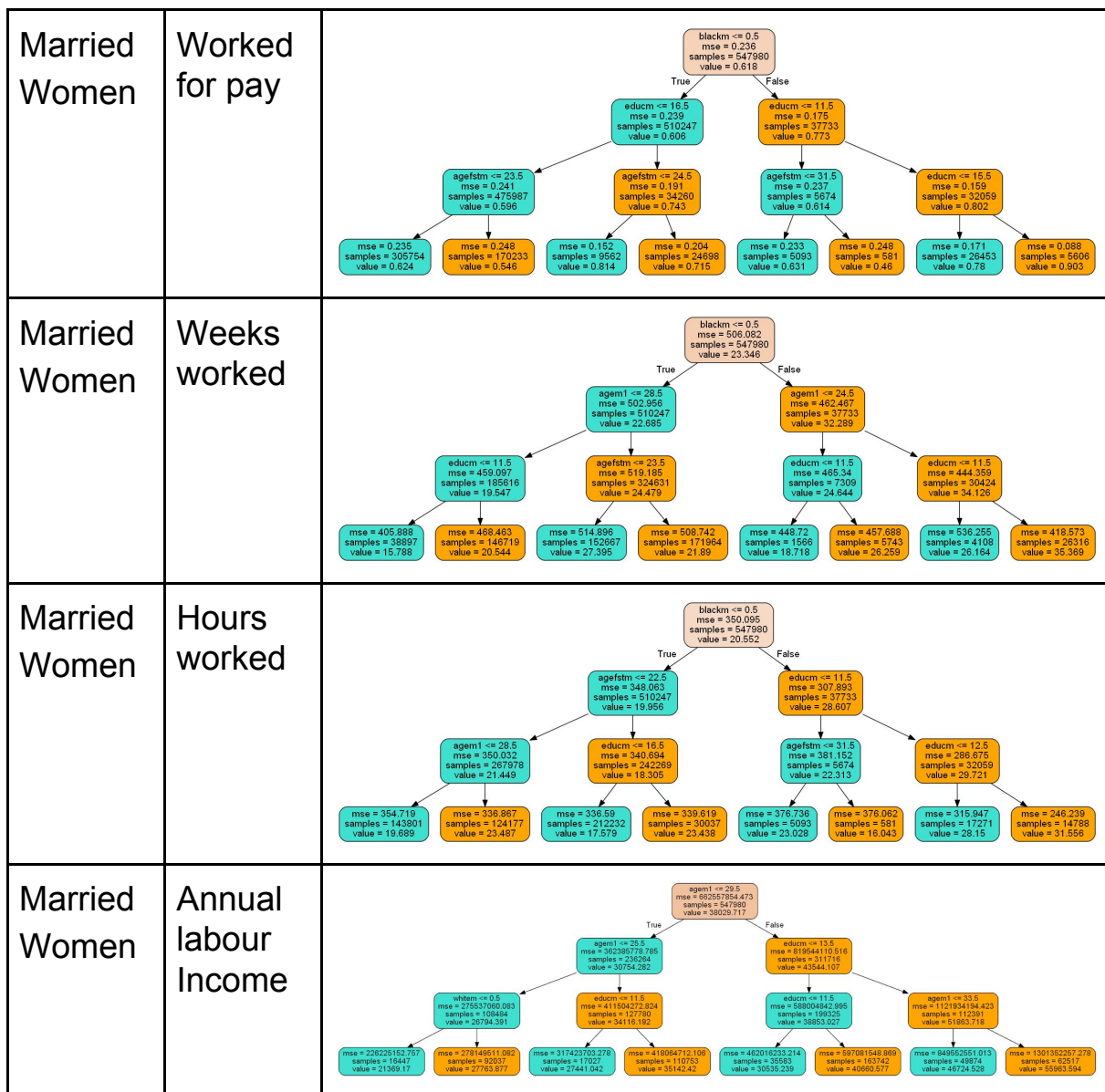
Married Women	Log family income	
---------------	-------------------	--

Treatment = 1

Husband	Worked for pay	
Husband	Weeks worked	
Husband	Hours worked	
Husband	Annual labour Income	



Treatment = 0



Married Women	Log family income	
---------------	-------------------	--

Treatment = 0

Husband	Worked for pay	
Husband	Weeks worked	
Husband	Hours worked	
Husband	Annual labour Income	

Husband	Log family income	<pre> graph TD Root["age1 <= 27.5 mse = 1.597 samples = 547680 value = 10.338"] Root -- True --> Node1["whitem <= 0.5 mse = 1.81 samples = 163494 value = 10.08"] Root -- False --> Node2["educ <= 11.5 mse = 1.548 samples = 378486 value = 10.453"] Node1 -- True --> Node3["age1 <= 24.5 mse = 2.659 samples = 24510 value = 9.761"] Node1 -- False --> Node4["educ <= 11.5 mse = 1.412 samples = 144984 value = 10.134"] Node2 -- True --> Node5["whitem <= 0.5 mse = 2.251 samples = 43524 value = 10.075"] Node2 -- False --> Node6["whitem <= 0.5 mse = 1.435 samples = 334962 value = 10.502"] Node3 -- True --> Node7["mse = 2.869 samples = 12621 value = 9.645"] Node3 -- False --> Node8["mse = 2.385 samples = 11889 value = 9.884"] Node4 -- True --> Node9["mse = 1.871 samples = 30842 value = 9.916"] Node4 -- False --> Node10["mse = 1.272 samples = 114142 value = 10.193"] Node5 -- True --> Node11["mse = 3.199 samples = 7673 value = 9.779"] Node5 -- False --> Node12["mse = 2.026 samples = 35851 value = 10.138"] Node6 -- True --> Node13["mse = 2.438 samples = 35205 value = 10.187"] Node6 -- False --> Node14["mse = 1.305 samples = 299757 value = 10.539"] </pre>
---------	-------------------	--

Transformed Outcome Trees

All Women	Worked for pay	<pre> graph TD Root["age1 <= 30.5 mse = 2.79 samples = 927267 value = -0.124"] Root -- True --> Node1["agefstm <= 18.5 mse = 1.902 samples = 420421 value = -0.397"] Root -- False --> Node2["agefstm <= 23.5 mse = 3.413 samples = 506846 value = 0.104"] Node1 -- True --> Node3["mse = 2.752 samples = 136529 value = -0.121"] Node1 -- False --> Node4["mse = 1.438 samples = 283892 value = -0.53"] Node2 -- True --> Node5["mse = 4.16 samples = 279796 value = 0.417"] Node2 -- False --> Node6["mse = 2.223 samples = 227050 value = -0.282"] </pre>
All Women	Weeks worked	<pre> graph TD Root["age1 <= 31.5 mse = 4613.634 samples = 927267 value = -5.975"] Root -- True --> Node1["agefstm <= 18.5 mse = 3009.197 samples = 479935 value = -14.587"] Root -- False --> Node2["agefstm <= 22.5 mse = 6170.068 samples = 447332 value = 3.265"] Node1 -- True --> Node3["mse = 4296.013 samples = 147698 value = -4.188"] Node1 -- False --> Node4["mse = 2367.692 samples = 332237 value = -19.21"] Node2 -- True --> Node5["mse = 8034.471 samples = 197905 value = 18.852"] Node2 -- False --> Node6["mse = 4345.059 samples = 249427 value = -9.103"] </pre>
All Women	Hours worked	<pre> graph TD Root["age1 <= 30.5 mse = 3505.586 samples = 927267 value = -4.949"] Root -- True --> Node1["agefstm <= 18.5 mse = 2488.488 samples = 420421 value = -13.919"] Root -- False --> Node2["agefstm <= 22.5 mse = 4227.153 samples = 506846 value = 2.491"] Node1 -- True --> Node3["mse = 3796.092 samples = 136529 value = -4.674"] Node1 -- False --> Node4["mse = 1798.76 samples = 283892 value = -18.365"] Node2 -- True --> Node5["mse = 5568.568 samples = 234418 value = 15.123"] Node2 -- False --> Node6["mse = 2817.473 samples = 272428 value = -8.378"] </pre>

All Women	Annual labour Income	<pre> graph TD Root["age1 <= 31.5 mse = 10018754143.714 samples = 927267 value = 3375.677"] Root -- True --> L1["agefstm <= 22.5 mse = 4873976458.956 samples = 479935 value = -9913.003"] Root -- False --> R1["agefstm <= 25.5 mse = 15145772611.814 samples = 447332 value = 17632.879"] L1 --> L2["mse = 4953695171.626 samples = 364907 value = -2525.674"] L1 --> L3["mse = 3898757582.816 samples = 115028 value = -33348.065"] R1 --> R2["mse = 15660491668.677 samples = 307295 value = 32617.4"] R1 --> R3["mse = 12442350612.041 samples = 140037 value = -15248.918"] </pre>
All Women	Log family income	<pre> graph TD Root["age1 <= 29.5 mse = 500.351 samples = 927267 value = -0.012"] Root -- True --> L1["agefstm <= 20.5 mse = 330.829 samples = 362913 value = -5.144"] Root -- False --> R1["agefstm <= 24.5 mse = 581.539 samples = 564354 value = 3.288"] L1 --> L2["mse = 406.787 samples = 223874 value = -2.312"] L1 --> L3["mse = 174.815 samples = 139039 value = -9.704"] R1 --> R2["mse = 631.235 samples = 364544 value = 7.629"] R1 --> R3["mse = 393.724 samples = 199810 value = -4.634"] </pre>

Married Women	Worked for pay	<pre> graph TD Root["age1 <= 30.5 mse = 2.785 samples = 762843 value = -0.079"] Root -- True --> L1["agefstm <= 20.5 mse = 1.892 samples = 332598 value = -0.367"] Root -- False --> R1["agefstm <= 23.5 mse = 3.362 samples = 430245 value = 0.143"] L1 --> L2["mse = 2.452 samples = 183144 value = -0.191"] L1 --> L3["mse = 1.122 samples = 149454 value = -0.581"] R1 --> R2["mse = 4.102 samples = 231815 value = 0.464"] R1 --> R3["mse = 2.237 samples = 198430 value = -0.231"] </pre>
Married Women	Weeks worked	<pre> graph TD Root["age1 <= 31.5 mse = 4501.302 samples = 762843 value = -3.967"] Root -- True --> L1["agefstm <= 20.5 mse = 2919.957 samples = 382339 value = -12.938"] Root -- False --> R1["agefstm <= 23.5 mse = 5928.146 samples = 380504 value = 5.047"] L1 --> L2["mse = 3771.558 samples = 201828 value = -6.373"] L1 --> L3["mse = 1865.722 samples = 180511 value = -20.278"] R1 --> R2["mse = 7459.49 samples = 198638 value = 18.008"] R1 --> R3["mse = 3871.724 samples = 181866 value = -9.108"] </pre>

Married Women	Hours worked	<pre> graph TD Root["agem1 <= 30.5 mse = 3383.93 samples = 762843 value = -3.248"] Root -- True --> L1["agefstm <= 18.5 mse = 2399.341 samples = 332598 value = -12.502"] Root -- False --> R1["agefstm <= 22.5 mse = 4027.674 samples = 430245 value = 3.906"] L1 --> L2["mse = 3836.996 samples = 95012 value = -2.476"] L1 --> L3["mse = 1768.136 samples = 237586 value = -16.512"] R1 --> R2["mse = 5329.911 samples = 192484 value = 16.551"] R1 --> R3["mse = 2739.178 samples = 237761 value = -6.331"] </pre>
Married Women	Annual labour Income	<pre> graph TD Root["agem1 <= 31.5 mse = 12175221648.125 samples = 762843 value = 4103.274"] Root -- True --> L1["agefstm <= 21.5 mse = 6086622819.078 samples = 382339 value = -12443.4"] Root -- False --> R1["agefstm <= 25.5 mse = 17741630684.988 samples = 380504 value = 20729.746"] L1 --> L2["mse = 6474054009.307 samples = 244074 value = -1111.583"] L1 --> L3["mse = 4775882410.737 samples = 138265 value = -32447.031"] R1 --> R2["mse = 18370690198.75 samples = 258610 value = 38757.836"] R1 --> R3["mse = 14254538665.38 samples = 121894 value = -17518.604"] </pre>
Married Women	Log family income	<pre> graph TD Root["agem1 <= 29.5 mse = 530.974 samples = 762843 value = 0.54"] Root -- True --> L1["agefstm <= 20.5 mse = 351.838 samples = 284938 value = -5.066"] Root -- False --> R1["agefstm <= 24.5 mse = 607.87 samples = 477905 value = 3.883"] L1 --> L2["mse = 445.76 samples = 164153 value = -1.717"] L1 --> L3["mse = 188.246 samples = 120785 value = -9.616"] R1 --> R2["mse = 657.596 samples = 302626 value = 8.602"] R1 --> R3["mse = 417.172 samples = 175279 value = -4.265"] </pre>

Husband	Worked for pay	<pre> graph TD Root["agem1 <= 29.5 mse = 5.042 samples = 762843 value = 0.049"] Root -- True --> L1["agefstm <= 20.5 mse = 3.511 samples = 284938 value = -0.501"] Root -- False --> R1["agefstm <= 24.5 mse = 5.666 samples = 477905 value = 0.377"] L1 --> L2["mse = 4.468 samples = 164153 value = -0.174"] L1 --> L3["mse = 1.868 samples = 120785 value = -0.946"] R1 --> R2["mse = 6.182 samples = 302626 value = 0.828"] R1 --> R3["mse = 3.82 samples = 175279 value = -0.4"] </pre>
---------	----------------	---

Husband	Weeks worked	<pre> graph TD Node0["agem1 <= 30.5 mse = 12304.462 samples = 762843 value = 2.274"] Node1["agefstm <= 20.5 mse = 8821.101 samples = 332598 value = -21.311"] Node2["agefstm <= 24.5 mse = 14234.775 samples = 430245 value = 20.507"] Node3["mse = 10932.528 samples = 183144 value = -4.139"] Node4["mse = 5429.545 samples = 149454 value = -42.355"] Node5["mse = 15456.975 samples = 267154 value = 43.438"] Node6["mse = 9960.482 samples = 163091 value = -17.055"] Node0 -- True --> Node1 Node0 -- False --> Node2 Node1 --> Node3 Node1 --> Node4 Node2 --> Node5 Node2 --> Node6 </pre>
Husband	Hours worked	<pre> graph TD Node0["agem1 <= 29.5 mse = 10537.682 samples = 762843 value = 2.754"] Node1["agefstm <= 20.5 mse = 7185.721 samples = 284938 value = -21.753"] Node2["agefstm <= 24.5 mse = 11964.6 samples = 477905 value = 17.366"] Node3["mse = 9004.004 samples = 164153 value = -7.378"] Node4["mse = 4052.065 samples = 120785 value = -41.29"] Node5["mse = 13117.602 samples = 302626 value = 37.353"] Node6["mse = 8093.291 samples = 175279 value = -17.143"] Node0 -- True --> Node1 Node0 -- False --> Node2 Node1 --> Node3 Node1 --> Node4 Node2 --> Node5 Node2 --> Node6 </pre>
Husband	Annual labour Income	<pre> graph TD Node0["agefstm <= 23.5 mse = 857908756.619 samples = 762843 value = -2285.166"] Node1["agem1 <= 32.5 mse = 855052459.421 samples = 507031 value = -217.378"] Node2["agem1 <= 36.5 mse = 838298054.59 samples = 255812 value = -6383.616"] Node3["mse = 557871621.25 samples = 343655 value = -3116.158"] Node4["mse = 1425306324.804 samples = 163376 value = 5880.091"] Node5["mse = 534032155.068 samples = 154245 value = -8726.561"] Node6["mse = 1279375597.05 samples = 101567 value = -2825.496"] Node0 -- True --> Node1 Node0 -- False --> Node2 Node1 --> Node3 Node1 --> Node4 Node2 --> Node5 Node2 --> Node6 </pre>
Husband	Log family income	<pre> graph TD Node0["agem1 <= 29.5 mse = 567.335 samples = 762843 value = 0.633"] Node1["agefstm <= 20.5 mse = 377.679 samples = 284938 value = -5.156"] Node2["agefstm <= 24.5 mse = 648.517 samples = 477905 value = 4.085"] Node3["mse = 477.159 samples = 164153 value = -1.69"] Node4["mse = 203.96 samples = 120785 value = -9.867"] Node5["mse = 700.547 samples = 302626 value = 8.945"] Node6["mse = 447.491 samples = 175279 value = -4.306"] Node0 -- True --> Node1 Node0 -- False --> Node2 Node1 --> Node3 Node1 --> Node4 Node2 --> Node5 Node2 --> Node6 </pre>