HOMEWORK #3

MARIA MOYA

1. PART I

C3.4

i.) ii.)

TABLE 1. minimum, maximum, and average values

variable	mean	sd	min	max
atndrte	81.71	17.05	6.25	100.00
priGPA	2.59	0.54	0.86	3.93
ACT	22.51	3.49	13.00	32.00

In order to estimate $atndrte = \beta_0 + \beta_1 priGPA + \beta_2 ACT + u$ we consider,

$\hat{atndrte} = 75.70 + 17.26 priGPA - 1.72 ACT$ where $N = 680$ and $R^2 = 0.29$. The coefficient, 75.70, is the predicted percent of classes attended given that the student has a GPA and ACT score of zero. Therefore, the coefficient is not particularly useful by itself.

iii.)A unit increase in priGPA results in a 17% increase in class attendance, on average.The sign of the coefficient is consistent with the logic that those who are more incline to increase their GPA are also more likely to attend class.

I anticipated a positive correlation between ACT and class attendance. One would expect, the higher an individual scores on the ACT the more incline that student is to attend class. However, the regression yields an unexpected negative correlation between ACT scores and class attendance.

iv.) If $priGPA = 3.65$ and $ACT = 20$ then using the equation from ii.) we get:

$104.30 = 75.7 + 17.26 \cdot (3.65) 1.72 \cdot (20)$. Thus a student with a 3.65 GPA and an ACT score of 20 will attend class 100% on average (I say 100% of the time because you can't show up for 104%).

There are two students that fit this criteria and only one with the exact values.That particular student had an class attendance rate of 87.5%.

v.) In order to find the predicted difference in their attendance rates, we must calculate the average attendance rate for each student (using the same method in iv.))and subtract them.

If Student A has $priGPA = 3.1$ and $ACT = 21$,

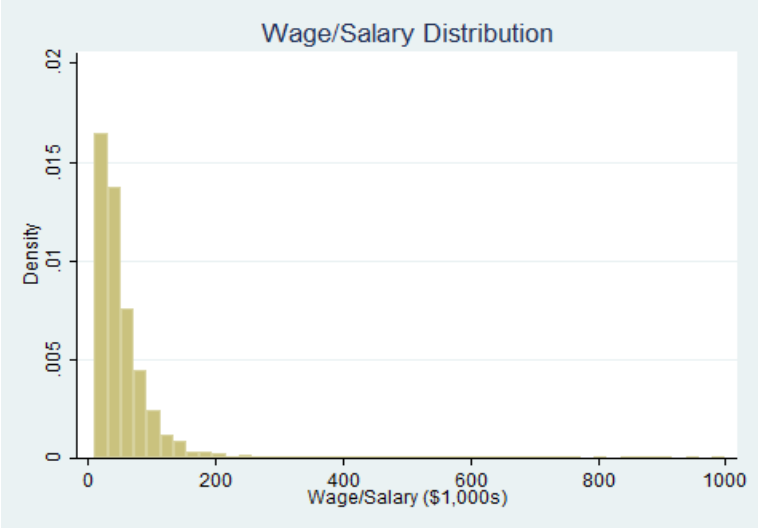
then $93.09 = 75.7 + 17.26 \cdot (3.1) 1.72 \cdot (21)$

Moreover if Student B has $priGPA = 2.1$ and $ACT = 26$

then $67.23 = 75.7 + 17.26 \cdot (2.1) 1.72 \cdot (26)$

Therefore, the predicted difference in their attendance rates is $93.09 - 67.23 = 25.86\%$

FIGURE 1. Wsal in 1000's



2. PART II

1.) This study includes March Supplement to the 2015 Current Population Survey (CPS) data which conducts a monthly survey to 54,000 household across the 50 states. These households are selected at random which include civilian non-institutional population of the United States living in housing units and members of the Armed Forces living in civilian housing units on a military base or in a household not on a military base.

2.)

a.)The calculation can be found under the comment:

$/ * \#2 percent female, median number of years of education....$

$display per\_female$

46.852737

b.)

TABLE 2. Median educ

Variable	Obs	Percentile	Centile	$[95\% Conf. Interval$
educ	69902	50	41	41 , 41

c.)

TABLE 3. Median educ

Variable	Obs	Percentile	Centile	$[95\% Conf. Interval$
wsal	69902	25	26	25, 26
		50	42	41, 42
		75	67	66, 68

3.)

4.)

5.) b.)for col, we would expect on average,as educational attainment increases,

FIGURE 2. Log of wsal

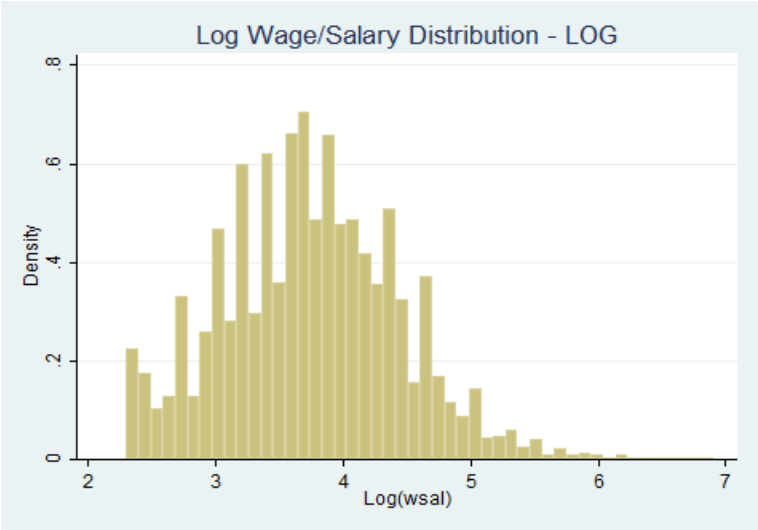


TABLE 4. TABLE 1:Summary Statistics

variable	mean	sd	min	max
wsal	55.57	54.97	10	1000
age	43.23	10.75	25	64
hs	0.26	0.44	0	1
col	0.24	0.43	0	1
white	0.79	0.41	0	1
metro_status	1.19	0.42	1	3
married	0.65	0.48	0	1
separated	0.02	0.15	0	1
divorced	0.11	0.32	0	1
female	0.47	0.5	0	1
nkids	0.93	1.15	0	12
nper	3.04	1.59	1	14
northeast	0.17	0.38	0	1
south	0.35	0.48	0	1
midwest	0.21	0.4	0	1
west	0.27	0.45	0	1

the log of wages should increase. Therefore we expect a positive coefficient.As far as the south variable, we would expect a negative coefficient. Cost of living in the South tend to be substantially lower than other parts of the US. For nper, as the number of people in the household increase we would have to pay more to sustain that household(this study does not include causal effects). Thus we would expect a positive coefficient. However the sign for nper is negative which may be due to our inability to account for causal effects on number of kids per household

c.)On average, holding all else constant,the difference in wages for an household with four children versus one child is $100 \cdot [e^{\beta_2} \cdot 3 - 1] = 100 \cdot [e^{.075 \cdot 3} - 1] = 25.23\%$

6.)

TABLE 5. Model 1

ln_wsal	Coef.	Std. Err.	t	$P > \|t\|$	95%*Conf.Interval*
age	.0079679	.0002514	31.69	0.0000	0.0074751, 0.0084607
hs	-.2514825	.0059343	-42.38	0.0000	-0.2631138, -0.2398512
col	.2277931	.006022	37.83	0.0000	0.2159901, 0.2395961
white	.0430472	.0060771	7.08	0.0000	0.0311362, 0.0549583
metro_status	-.1232981	.0059149	-20.85	0.0000	-0.1348913, -0.1117048
married	.2628639	.0070594	37.24	0.0000	0.2490275, 0.2767003
separated	-.0346763	.0168189	-2.06	0.039	-0.0676413, -0.0017113
divorced	.0798256	.0093869	8.5	0.0000	0.0614274, 0.0982238
female	-.3145733	.0049091	-64.08	0.0000	-0.3241951, -0.3049514
nkids	.0759714	.0036965	20.55	0.0000	0.0687263, 0.0832165
nper	-.068349	.0027973	-24.43	0.0000	-0.0738318, -0.0628662
south	-.0758764	.0071661	-10.59	0.0000	-0.0899218, -0.0618309
midwest	-.0828963	.007948	-10.43	0.0000	-0.0984744, -0.0673183
west	-.0744169	.0074619	-9.97	0.0000	-0.0890421, -0.0597916
_cons	3.695377	.0155447	237.73	0.0000	3.664909, 3.725844

a.)Since nkids is a discrete variable not continuous, it makes sense to consider subcases

b.)

TABLE 6. Model 2

ln_wsal	Coef.	Std. Err.	t	$P > \|t\|$	95%*Conf.Interval*
age	0.0084061	0.0002541	33.08	0	0.0079081, 0.0089042
hs	-0.2498241	0.0059289	-42.14	0	-0.2614448, -0.2382033
col	0.2267968	0.0060155	37.7	0	0.2150064, 0.2385871
white	0.0415596	0.0060712	6.85	0	0.0296601, 0.0534591
metro_status	-0.1223691	0.0059078	-20.71	0	-0.1339483, -0.1107899
married	0.2444743	0.0071723	34.09	0	0.2304167, 0.258532
separated	-0.0429796	0.0168186	-2.56	0.011	-0.075944, -0.0100153
divorced	0.0708293	0.0094158	7.52	0	0.0523743, 0.0892843
female	-0.3172384	0.0049109	-64.6	0	-0.3268638, -0.307613
nkids1	0.1077791	0.0074242	14.52	0	0.0932276, 0.1223306
nkids2	0.2125239	0.0090277	23.54	0	0.1948296, 0.2302183
nkids3	0.2265106	0.0131296	17.25	0	0.2007767, 0.2522446
nkids4	0.2129555	0.0198188	10.75	0	0.1741107, 0.2518003
nper	-0.0661158	0.0027568	-23.98	0	-0.0715191, -0.0607126
south	-0.0738726	0.0071594	-10.32	0	-0.087905, -0.0598402
midwest	-0.0799234	0.0079416	-10.06	0	-0.0954889, -0.064358
west	-0.0710938	0.007458	-9.53	0	-0.0857114, -0.0564761
$_{c}$ons	3.667016	0.0156466	234.37	0	3.636349, 3.697684

The coefficient for nkids4 is .2126868. This implies, on average, a household with four children will earn 21.3% more than a household with no children. Since nkids is discrete, Model 2 is preferred to Model 1. Nonetheless, neither regression controls for causal effects

c.)

3. PART III

1.) Recent studies have demonstrated the correlation between proximity of fast food restaurants and obesity rates however they fail address control-test that capture the effect of increase calories due to exercises, dietary habits that lead to a preference/increased demand for these restaurants. These are not casual. The author argues if dietary habits play a predominate factor, eliminating restaurants within a given proximity will lead people to converge to areas that will best satisfy their caloric intake

2.) The author focuses on the impact of opening restaurants near interstate highways among rural communities with high obesity rates. Restaurants that are built on interstate highways are not a result of increased demand by local residents. Therefore, since these restaurants are established absent of local demand this allows us capture the effect of opening a restaurant

3.) Obesity rates are due to the demand for high calorie food.If we were to completely eliminate restaurants, consumers would simply find substitutes to satisfy their caloric intake. The problem is not the abundance of restaurants rather the demand for caloric intake that lead to increased obesity

4.) For starters there is a selection bias among consumers who visit restaurants. Individuals who are more incline to eat at restaurants will most likely consume more calories. Furthermore, consumers who eat at restaurants smooth out their calorie intake, which decreases the observed effect calorie intake contributed solely via restaurants

5.)

Sources:phone surveys of individuals conducted by state departments.

unit of observation:the individual

sample size:2,608,980

geographic area:ZIP codes from AR, CO, IO, KS, ME, MO, ND, NE, OK, UT, and VT that are located less than 10 miles from an Interstate Highway and more than 30 miles from an urban area, and have a population density of less than 80 persons per square mile.

4. PART IV

1.) Voucher programs: Low income families are less incline to take advantage of scholarship voucher programs, despite larger returns relative higher income families. I want to investigate the different effects that lead up to the asymmetric information among low income families and different alternatives to encourage participation.

2.) The talk of the Democratic elections have been to increase minimum wage. However a flat $15 increase would have negative effects in places such as Gainesville,FL(and the vast majoirity of the US. I think they gathered that number based on huge metropolitan areas such as Boston and NY. Most areas of the United States obviously do not share the population density or cost of living as NY or Boston). For starters, if a candidate wants to increase minimum wage they should at least adjust it by a percentage relative to the local cost of living. I would like to create an index that determines an optimal increase to minimum wage and measure its effects