Poverty and Educational Attainment in St. Cloud & Central Minnesota

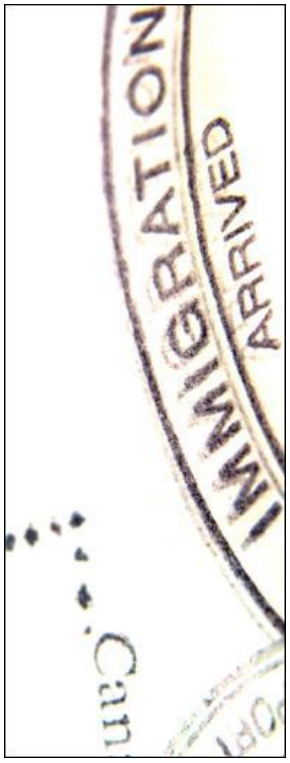
FALL2019

An analysis of Household Poverty and Education

Hamza Junaid & Claude Lee

Econ 485 Final Project

Contents



Executive Summary

Additional educational attainment is often believed to increase your future income, and such income incensement may change your poverty status. The analysis of differences in poverty status based on educational attainment was carried among residents of St. Cloud area and Central Minnesota. A Linear Probability Model for a poverty status was constructed using educational attainment, sex, race, age, number of children in a household, employment status, and healthcare coverage status.

Based on our analysis, for both in St. Cloud area and Central Minnesota:

- Higher education attainment population has less in-poverty proportion than lower education attainment population.
- Non-white and female population has more in-poverty proportion than white and male population, independently.
- More educational attainment leads to lower likelihood of being in poverty.
- Being non-white and female raise the likelihood of being in poverty.

Introduction

Poverty by definition is the state of living in extreme conditions where individuals or households are unable to fulfill their basic needs. Poverty is most often measured in monetary terms, and US Census measures poverty by capturing household income and comparing it to the Federal Poverty Line. Educational attainment by individuals is one of the important factors in the determination of poverty, since it impacts the ability of that individual to gain higher income. We study this exact relationship in the report and show how significant the impact of this relationship is.

Does Educational Attainment help in reducing the likelihood of being in poverty?

The analysis conducted shows the effect of educational attainment on poverty level in St. Cloud area and Central Minnesota. A linear probability model was fitted, using several economic and demographic indicators and poverty levels were predicted based on these indicators.

We took a holistic approach in basing our analysis on the human capital model that views individuals as economic units and can that they can be treated as a commodity. The human capital model is the background of our regression, variables for which were picked based on various reports.

As per the analysis, educational attainment does help in reducing the likelihood of being in poverty. At the same time, non-white and female population faces higher likelihood of being in poverty compared to white and male population.

The subsequent sections of this report give a broader view of the methods, analysis and regressions conducted for this project. Each section contains a detailed insight of the different parts of this project and how it all comes together as one big moving part that helps us present our findings.

Background and Literature Review

As per the human capital model discussed in the paper “Education's Effect on Poverty: The Role of Migration” (Weber et al., 2007) , people with more years of education will have more knowledge and more ability to work, which would allow them to perform better in workplace. The better performance in workplace could be translated into higher productivity, and workers with higher productivity are more competitive in labor market, leading to higher incomes for them. Since poverty level is a relative measure of income, as the household income increases, the poverty level would decrease mathematically. Therefore, more years of education would lead to lower poverty level in a household. This is a crucial finding that helps us establish the basis of the relationship and work on building a model around it.

Due to gender and racial discrimination, both historical and current, with consequent effects on their environment, people in minority groups are less competitive workers compared to people in majority groups, pushing the minority groups into higher poverty level. This has been discussed in the paper “Race, Ethnicity, and the Gender-Poverty Gap” (Reeves et al., 2016; Lu et al., 2002; Chaudhuri, 2018; Creamer et al., 2018). Since higher education has costs, being poverty limits the years of education that those minorities can have, and eventually they will end up having even higher poverty level. Along with gender and race, healthcare coverage, employment status, number of children, and age can also affect the poverty level (Reeves et al., 2016)

For poverty, Federal Poverty Level (FPL) is used to define poverty status. Federal poverty level is defined by ratio of income to poverty threshold in percentage terms. The status is represented as In-Poverty (IP) for FPL < 150% and Not-In-Poverty (NP) for FPL ≥ 150% (Reeves et al., 2016). For education, 3 categories were made to represent the latest education attainment. These include High School Dropouts (HSD), High School Graduates (HS), and some college (SC) (See appendix A).

A Snapshot of the population

5%: 5-in-100 national random sample of the population

Data

The data used in the analysis is extracted from Integrated Public Use Microdata Series (IPUMS). This data is consisted of American Community Survey (ACS) from 2013 to 2017 with 5% density, meaning randomly selected 5% of entire population constitute the data. Each observation is weighted by *PERWT* variable, which estimates how many persons in the U.S. population are represented by the given person in the row. The data is restricted to have only samples within working age (24-65), while excluding observations that are lack of some of the information that are necessary in the analysis.

St. Cloud Area

Since St. Cloud is not identified as a City in IPUMS dataset, Stearns County is used as a proxy for St. Cloud area. Stearns County data is filtered through setting *COUNTYICP=1450*, where *COUNTYICP* uses Inter-University Consortium for Political and Social Research (ICPSR) coding to separate counties within states. Sample size for this data is 4,359, whereas the population estimate using *PERWT* variable is 78,551. However, US Census Bureau reported that the population of Stearns County is 155,300. Therefore, the weighted counts may not reflect the true population well.

Central Minnesota

Central Minnesota is a geographically central area of the state of Minnesota with no specific boundaries. In this analysis, Central Minnesota is approximated by setting *PUMA=600* (Pine, Kanabec, Mille Lacs, Isanti, Chisago)*, 900* (Stearns)*, 1000* (Sherburne, Benton)*, 1800* (Wright)*, 1900* (Kandiyohi, Meeker, McLeod, Renville, Sibley)*,* where *PUMA* uses Public Use Microdata Area coding to separate boundaries of county groups. Sample size for this data is 22,735, whereas the population estimate using *PERWT* variable is 379,157. US Census Bureau reported this as 924,982, and this has the same issue as St. Cloud area population estimate.

Summary Statistics

Table 1. Summary Statistics for St. Cloud Area and Central Minnesota

Subgroup	St. Cloud area					Central Minnesota				
	In Poverty		Not In Poverty		χ^2Test	In Poverty		Not In Poverty		χ^2Test
	Prop.	WC**	Prop.	WC**	p-value	Prop.	WC**	Prop.	WC**	p-value
HSD	45.67	1878	54.33	2234		43.94	7941	56.06	10132	
HS	16.61	4238	83.39	21273	< 0.0001 ***	14.98	21435	85.02	121616	< 0.0001 ***
SC	12.19	5963	87.81	42965		8.22	17927	91.78	200107	
Male	12.99	5114	87.01	34244	< 0.0001 ***	10.98	21112	89.02	171191	< 0.0001 ***
Female	17.77	6965	82.23	32228		14.02	26191	85.98	160664	
White	13.29	9614	86.71	62718	< 0.0001 ***	11.14	39784	88.86	317259	< 0.0001 ***
Non-White	39.64	2465	60.36	3754		34	7519	66	14596	
Age*	40	12079	45	66472	< 0.0001 ***	42	47303	45	331855	< 0.0001 ***
No Children	15.02	6264	84.98	35451	0.0006 ***	12.8	24307	87.2	165649	< 0.0001 ***
Have Children	15.79	5815	84.21	31021		12.15	22996	87.85	166206	
Employed	10.99	7275	89.01	58949	< 0.0001 ***	8.21	25675	91.79	287165	< 0.0001 ***
Not Employed	38.97	4804	61.03	7523		32.61	21628	67.39	44690	
Healthcare Covered	13.47	9953	86.53	63945	< 0.0001 ***	11.46	40756	88.54	314890	< 0.0001 ***
NC Not Covered	45.69	2126	54.31	2527		27.85	6547	72.15	16965	
Total Proportion	15.38	12079	84.62	66472		12.48	47303	87.52	331855	

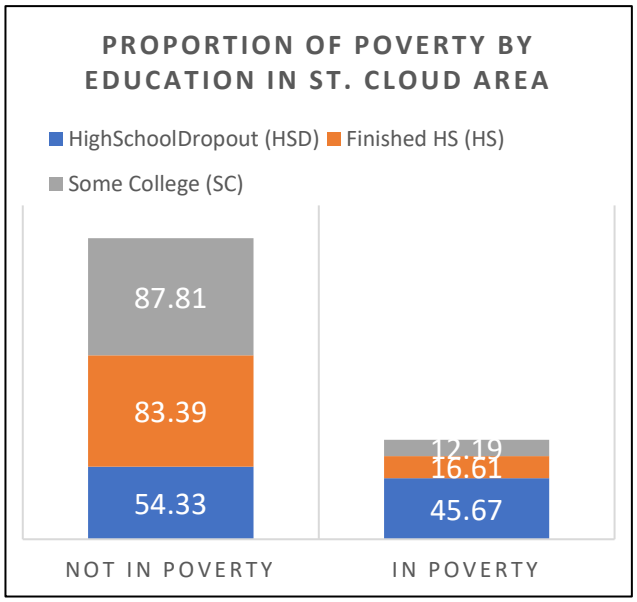
*For continuous variable *Age*, average is calculated in proportion cell, t-test is used for χ^2Test.
**WC: Weighted Counts or Population estimates using *PERWT* variable.

The table represents the proportion of in-poverty and not-in-poverty within each subgroup (used row percentage). Each χ^2 test tests whether subgroups within each variable have different proportion of poverty. All χ^2 test results show the evidence that each variable differs the poverty proportion as their subgroup changes for both regions. The detailed discussion about summary statistics of educational attainment is stated in the next section.

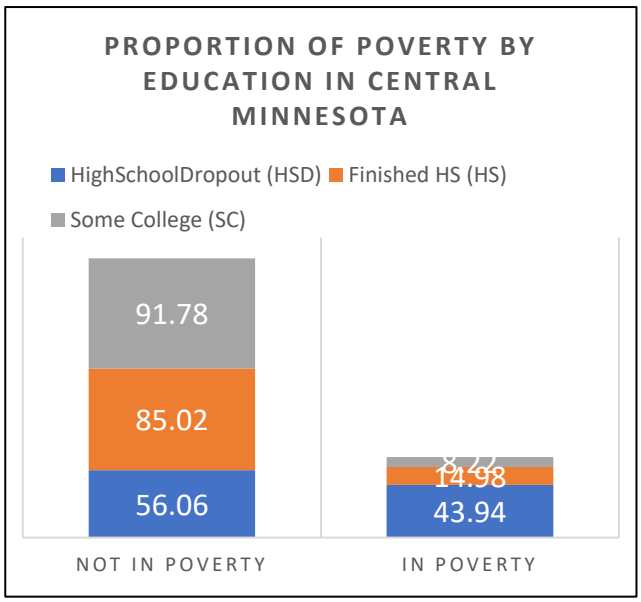
Except for educational attainment, the variables sex, race (white/non-white), employment status, and healthcare coverage also show noticeable proportion differences in poverty status. However, having or not having children does not have much difference in the proportion within the actual unit even when χ^2 test result state differently.

Does Educational Attainment Affect Poverty Status?

Graph 1. St. Cloud



Graph 2. Central Minnesota



Graph 1 and 2 visually show that as education attainment increases, the less people are in poverty. This result is consistent with the χ^2 test. Additionally, female and non-white population have higher 'in poverty proportion' compared to male and white population, respectively. However, χ^2 test compares one variable to poverty at a time. This only means correlation between the two variable, and correlation does not mean causation.

Focus: Partial Effects on Poverty

From Table 1 and Graph 1&2, general trend of educational attainment on poverty status has been set, along with sex and race. Now, econometric analysis is needed to see the true effect of educational attainment, sex, and race on poverty proportion, while controlling for other variables that also affects poverty status.

Econometric Analysis: *Linear Probability Model*

Linear Probability Model (LPM) is a simple OLS linear regression with indicator variable as a response. Since the response will be an average of 0 and 1 for observations specified by dependent variables, the model output becomes a probability of having such indication. Here, "*Poverty*" becomes an indicator of being in poverty as our response variable. From previous literature reviews and analyzing sample size, reference group is established as "high school dropout white males without children, employed with health insurance." Each independent variable is defined as table 2 below. The reference group is marked as red in each variable.

Table 2. Indicator Variables

POVERTY		EDUC		FEMALE		NWHITE		YCHILD		NEMP		NHCOV	
0	NP	**0**	**HSD**	**0**	**Male**	**0**	**White**	**0**	**No children**	**0**	**Employed**	**0**	**Healthcare Covered**
1	IP	1	HS	1	Female	1	Non-White	1	Have children	1	Not employed	1	No HC Cov.
		2	SC										

EDUC: latest educational attainment
FEMALE: is female (in sex)
NWHITE: is non-white (racially)
YCHILD: has children
NEMP: not employed
NHCOV: healthcare is not covered

Econometric Model

$$\boldsymbol{POVERTY} = \beta_0 + \beta_{1,l}\boldsymbol{EDUC_l} + \beta_2\boldsymbol{FEMALE} + \beta_3\boldsymbol{NWHITE} + \beta_4\,\boldsymbol{Age} + \beta_5\boldsymbol{YCHILD} + \beta_6\boldsymbol{NEMP} + \beta_7\boldsymbol{NHCOV} + \varepsilon$$

- $\boldsymbol{EDUC_l}$ represents each educational attainment level for $l = 0,1,2$, and $\beta_{1,l}$ represents corresponding coefficients for each educational attainment level.

Partial Effects of Educational Attainment on Poverty

Table 3. LPM Parameter Estimates

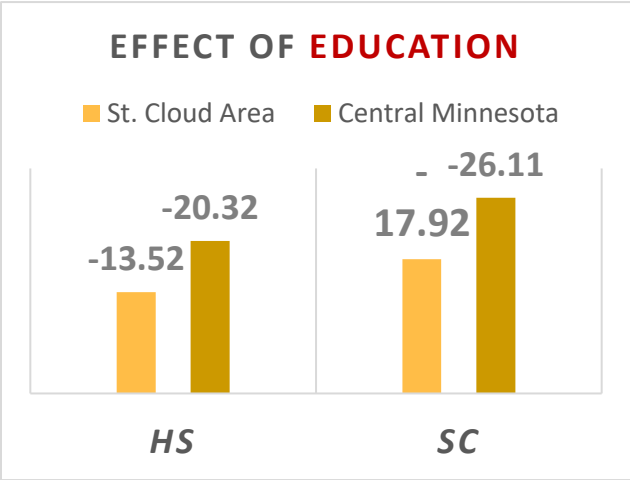
Estimated Effects of Each Group on Probability of Being in Poverty				
Parameter %point unit	St. Cloud Area		Central Minnesota	
	Estimate	(Sig) Std Error	Estimate	(Sig) Std Error
Intercept	43.26	*** 3.20	43.96	*** 1.32
EDUC – HS	**-13.52**	*** 2.47	**-20.32**	*** 1.01
EDUC – SC	**-17.92**	*** 2.41	**-26.11**	*** 1.00
FEMALE	**3.58**	*** 1.01	**2.44**	*** 0.41
NWHITE	**13.87**	*** 1.96	**14.19**	*** 0.89
AGE	-0.47	*** 0.04	-0.34	*** 0.02
YCHILD	**0.21**	1.02	**-0.45**	0.42
NEMP	26.00	*** 1.42	22.67	*** 0.55
NHCOV	23.51	*** 2.16	8.96	*** 0.85

** Significance level from t-test is marked with Std Error.

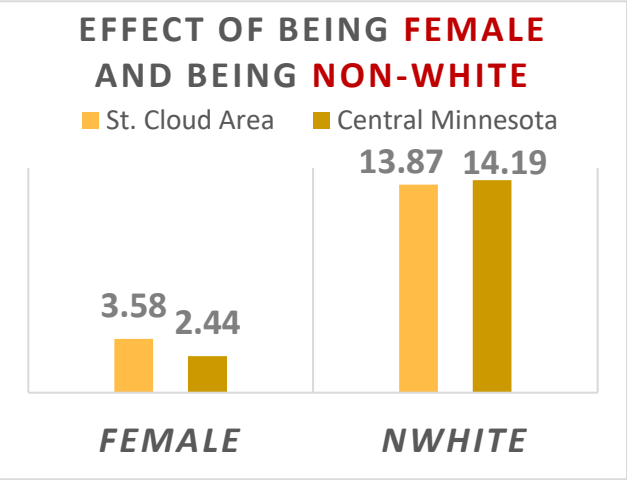
Since the response variable, *Poverty,* is an indicator variable, the model would generally return values in between 0 and 1. In this table, the unit is converted to percentage point, accordingly *Poverty* would be returned in between 0 and 100, which can be interpreted as "*probability*". Therefore, the coefficients represent how the each indicator variable affects the "*probability of being in poverty*," in other words, "*likelihood of being in poverty*".

Partial Effects of Educational Attainment on Poverty

Graph 1. Effect of Education



Graph 2. Effect of Being Female and Being Non-White



As a result, the LPM estimates show negative signs for both finishing high school degree and having some college education, and negative coefficients mean lower predicted value of *Poverty*. Therefore, obtaining extra education decreases the likelihood of being in poverty. The effect size is greater for Central Minnesota compared to St. Cloud area as shown in Graph 1. On the other hand, estimates for being female and being non-white are positive, meaning being female and being non-white increase the likelihood of being in poverty. The effect size for this indicators are relatively similar in both regions as shown in Graph 2.

The results for educational attainment, sex, and race are consistent with the literature reviews and economic background set up for the report. Being employed and having healthcare covered also change the likelihood of being in poverty as expected. However, notice that having or not having children does not change likelihood of being poverty much. In St. Cloud, having children raises the likelihood of being in poverty by 0.2 percentage point, whereas in Central Minnesota, it reduces by 0.5 percentage point. The standard error for this indicator also suggests that this coefficient is not significant.

Concluding remarks

From previous Linear Probability Model, the estimates have established that extra education decreases probability of being in poverty. However, the model may not be the best estimates for analyzing relationship between education and poverty due to following reasons:

- **Selection bias**

Why these highly educated people choose to stay in St. Cloud area or Central Minnesota? The factors that made them to stay in these regions have relationship with their income as well, such as their majors and types of the jobs they get etc. Without accounting for these factors, the estimates from LPM would have omitted variable bias.

- **Reverse Causality**

Being in poverty also effects the education level. If you're poor, the level of education that you get might also be lower like if you can't afford to get higher education. Therefore, partial effect of education attainment may not accurately reflect the true partial effect.

Additionally, model validation and improvements are needed to verify that the model represents unbiased coefficients. This step includes residual diagnostics for MLR assumption verification, comparing other models such as using Logistics model, re-selection of variables such as removing *YCHILD* variable since it is not significant in LPM estimates (marked as red in table 3).

Furthermore, research can be expanded to scrutinize each variable and effects more specifically, such as expanding *NWHITE* as multiple races, or adding interaction term between education and sex to see how much of education affect changes by sex.

References

Bruce Weber, A. M. (2007). Education's Effect on Poverty: The Role of Migration. *Agricultural Economics* .

Chaudhuri, S. (2018). *Education and the Gender Gap.* Minnesota Department of Employment and Economic Development .

Jr., A. C. (2006). *The More The Poorer: Why Large Family Size Causes Poverty.* Philippine Institute for Development Studies .

Richard. Reeves, E. R. (2016). *Five Evils: Multidimensional Poverty and Race in America.* The Brookings Institution .

(2014). *Status of Women & Girls in Minnesota.* University of Minnesota Humphrey School's Center on Women & Public Policy .

(2016). *The Economic Status of Minnesotans: A Chartbook With Data for 17 Cultural Groups.* Minnesota State Demographic Center .

Yuval Elmelech, H.-H. L. (2002). Race, Ethnicity, and the Gender-Poverty Gap. *Columbia University*.

Appendix

A. Key Variable Categorization

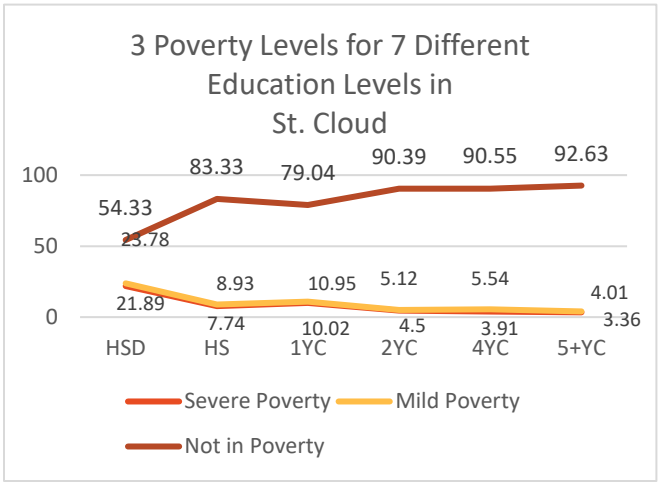
Table A1. Initial Categories for Poverty

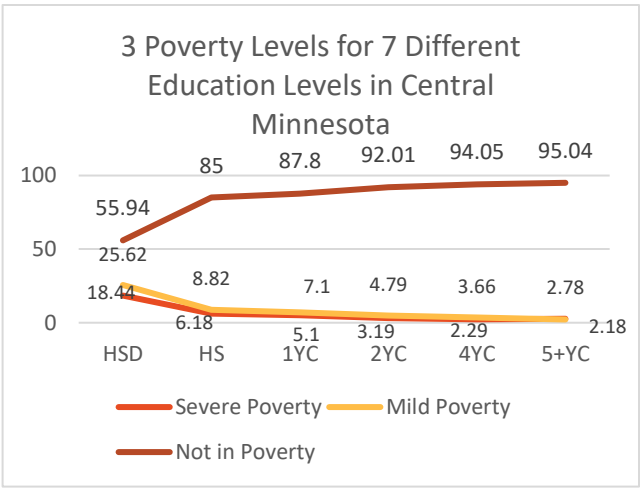
Poverty Level	Poverty Status *(POVERTY)*
FPL ≤ 75%	In Severe Poverty
75% ≤ FPL < 150%	In Mild Poverty
FPL ≥ 150%	Not in poverty

Table A2. Initial Categories for Education

Highest Educational Attainment	Abbr.
High School Dropouts	HSD
High School Graduates	HS
1 Year in College	1YC
2 Years in College	2YC
3 Years in College	3YC
4 Years in College	4YC
5+ Years in College	5+YC

Graph A1. 3-Level Poverty 7-Level Education





**3YC sample size is zero for both samples

Initially, the analysis began with 3 levels in poverty level and 7 levels in education attainment as Table A1. And Table A2. However, there is only a slight difference between severe poverty and mild poverty for different education attainment. Also, after HS, the poverty trend lines are similar for any additional college education. Therefore, for simplicity, poverty variable was re-categorized as binary by combining severe and mild poverty, and education variable was as HSD, HS, and SC by combining all college education into one.

B. SAS Code

This report uses SAS Studio for outputting tables and estimating LPM models. For graphs, excel is used. Codes for creating each table and estimates are given below.

- Categorizing Variables

```
data work.POVBin;
	set ECON485.USA_00016;

	*Education level into 3;
	select;
		when (0 <=EDUC <=5) EDUCL=0; *HSD;
		when (EDUC = 6) EDUCL=1;     *HS;
		otherwise EDUCL=2;           *SC;
	end;

	*Poverty into binary;
	select;
		when (0 < poverty < 150) POV=1;
		when (poverty = 0) POV = "";
		otherwise POV=0;
	end;

	*Female indicator;
	select;
		when (SEX = 1) FEMALE = 0;
		when (SEX = 0) FEMALE = "";
		otherwise FEMALE = 1;
	end;

	*Non-White indicator;
	select;
		when (RACE = 1) NWHITE = 0;
		when (RACE = 0) NWHITE = "";
		otherwise NWHITE = 1;
	end;

	*Child indicator;
	select;
		when (NCHILD >= 1) YCHILD = 1;
		when (NCHILD = 0) YCHILD = 0;
```

```
            otherwise YCHILD = "";
      end;

      *No-Employment indicator;
      select;
            when (EMPSTAT = 1) NEMP = 0;
            when (EMPSTAT = 0) NEMP = "";
            otherwise NEMP = 1;
      end;

      *No-Healthcare indicator;
      select;
            when (HCOVANY = 1) NHCOV = 1;
            otherwise NHCOV = 0;
      end;

      if POV = "" OR FEMALE = "" OR NWHITE = "" OR YCHILD = "" OR NEMP = "" THEN
DELETE;

run;
```

- Sample Size of Each Regions

```
proc freq data=work.POVBin(where = (countyicp in (1450)));
      tables  POV /  nopercent nocol plots=none;
      weight perwt;
run;
proc freq data=work.POVBin(where = (puma in (600 900 1000 1800 1900)));
      tables  POV /  nopercent nocol plots=none;
      weight perwt;
run;
```

- Summary Statistics (Table 1)

```
**************
Stearns County
**************;


*****
POVERTY
*****;

proc freq data=work.POVBin(where = (countyicp in (1450)));
      tables POV / plots=none;
      weight perwt;
      title "Poverty Level Proportion (Estimated Weighted, Stearns County, Aged 24-
65)";
run;
title;



*****
EDUC
*****;

proc freq data=work.POVBin(where = (countyicp in (1450)));
      tables  (EDUCL) *(POV) /  nopercent nocol nocum plots=none;
      weight perwt;
      title "Education Level Proportion in Poverty (Estimated Weighted, Stearns County,
Aged 24-65)";
run;
title;



*****
FEMALE
*****;

proc freq data=work.POVBin(where = (countyicp in (1450)));
      tables  (FEMALE) *(POV) /  nopercent nocol nocum plots=none;
      weight perwt;
      title "Sex Proportion in Poverty (Estimated Weighted, Stearns County, Aged 24-
65)";
run;
title;
```

```
*****
NWHITE
*****;

proc freq data=work.POVBin(where = (countyicp in (1450)));
       tables  (NWHITE) *(POV) /  nopercent nocol nocum plots=none;
       weight perwt;
       title "White/Non-White Proportion in Poverty (Estimated Weighted, Stearns County,
Aged 24-65)";
run;
title;



*****
AGE
*****;
proc means data=work.POVBin(where = (countyicp in (1450))) chartype mean std min max n
vardef=df;
       var AGE;
       class POV;
       weight perwt;
       title "Age Average in Poverty (Estimated Weighted, Stearns County, Aged 24-65)";
run;
title;



*****
YCHILD
*****;
proc freq data=work.POVBin(where = (countyicp in (1450)));
       tables  (YCHILD) *(POV) /  nopercent nocol nocum plots=none;
       weight perwt;
       title "Child-In-HH Proportion in Poverty (Estimated Weighted, Stearns County,
Aged 24-65)";
run;
title;



*****
NEMP
*****;
```

```
proc freq data=work.POVBin(where = (countyicp in (1450)));
      tables  (NEMP) *(POV) /  nopercent nocol nocum plots=none;
      weight perwt;
      title "Employment Status Proportion in Poverty (Estimated Weighted, Stearns
County, Aged 24-65)";
run;
title;



*****
NHCOV
*****;

proc freq data=work.POVBin(where = (countyicp in (1450)));
      tables  (NHCOV) *(POV) /  nopercent nocol nocum plots=none;
      weight perwt;
      title "Healthcare Coverage Status Proportion in Poverty (Estimated Weighted,
Stearns County, Aged 24-65)";
run;
title;




**************
Central Minnesota
**************;


*****
POVERTY
*****;

proc freq data=work.POVBin(where = (puma in (600 900 1000 1800 1900)));
      tables POV / plots=none;
      weight perwt;
      title "Poverty Level Proportion (Estimated Weighted, Central Minnesota, Aged 24-
65)";
run;
title;



*****
EDUC
```

```
*****;

proc freq data=work.POVBin(where = (puma in (600 900 1000 1800 1900)));
       tables  (EDUCL) *(POV) /  nopercent nocol nocum plots=none;
       weight perwt;
       title "Education Level Proportion in Poverty (Estimated Weighted, Central
Minnesota, Aged 24-65)";
run;
title;



*****
FEMALE
*****;

proc freq data=work.POVBin(where = (puma in (600 900 1000 1800 1900)));
       tables  (FEMALE) *(POV) /  nopercent nocol nocum plots=none;
       weight perwt;
       title "Sex Proportion in Poverty (Estimated Weighted, Central Minnesota, Aged 24-
65)";
run;
title;



*****
NWHITE
*****;

proc freq data=work.POVBin(where = (puma in (600 900 1000 1800 1900)));
       tables  (NWHITE) *(POV) /  nopercent nocol nocum plots=none;
       weight perwt;
       title "White/Non-White Proportion in Poverty (Estimated Weighted, Central
Minnesota, Aged 24-65)";
run;
title;



*****
AGE
*****;
proc means data=work.POVBin(where = (puma in (600 900 1000 1800 1900))) chartype mean
std min max n vardef=df;
       var AGE;
       class POV;
```

```
      weight perwt;
      title "Age Average in Poverty (Estimated Weighted, Central Minnesota, Aged 24-
65)";
run;
title;



*****
YCHILD
*****;
proc freq data=work.POVBin(where = (puma in (600 900 1000 1800 1900)));
      tables  (YCHILD) *(POV) /  nopercent nocol nocum plots=none;
      weight perwt;
      title "Child-In-HH Proportion in Poverty (Estimated Weighted, Central Minnesota,
Aged 24-65)";
run;
title;



*****
NEMP
*****;

proc freq data=work.POVBin(where = (puma in (600 900 1000 1800 1900)));
      tables  (NEMP) *(POV) /  nopercent nocol nocum plots=none;
      weight perwt;
      title "Employment Status Proportion in Poverty (Estimated Weighted, Central
Minnesota, Aged 24-65)";
run;
title;



*****
NHCOV
*****;

proc freq data=work.POVBin(where = (puma in (600 900 1000 1800 1900)));
      tables  (NHCOV) *(POV) /  nopercent nocol nocum plots=none;
      weight perwt;
      title "Healthcare Coverage Status Proportion in Poverty (Estimated Weighted,
Central Minnesota, Aged 24-65)";
run;
title;
```

- Linear Probability Model Estimates

```
*St. Cloud Area (Stearns County);
proc glmselect data=work.POVBin(where = (countyicp in (1450)))
outdesign(addinputvars)=Work.reg_design;
	class EDUCL(ref = "0") FEMALE(ref = "0") NWHITE(ref = "0") YCHILD(ref = "0")
NEMP(ref = "0") NHCOV(ref = "0") / param=ref;
	model POV=EDUCL FEMALE NWHITE AGE YCHILD NEMP NHCOV / showpvalues selection=none;
	weight PERWT;
	title "Linear Probability Model for Poverty in St. Cloud Area";
run;
title;
proc delete data=Work.reg_design;
run;

*Central Minnesota;
proc glmselect data=work.POVBin(where = (puma in (600 900 1000 1800 1900)))
outdesign(addinputvars)=Work.reg_design;
	class EDUCL(ref = "0") FEMALE(ref = "0") NWHITE(ref = "0") YCHILD(ref = "0")
NEMP(ref = "0") NHCOV(ref = "0") / param=ref;
	model POV=EDUCL FEMALE NWHITE AGE YCHILD NEMP NHCOV / showpvalues selection=none;
	weight PERWT;
	title "Linear Probability Model for Poverty in Central Minnesota";
run;
title;
proc delete data=Work.reg_design;
run;
```