Tim Donley

Econ 562

12/7/16

**On Socioeconomics  
  
  
Abstract**

Socioeconomics is a popular buzzword in the news with new social justice movements appearing daily across the world. Socioeconomic standing however is a very important and useful metric. Sociologists can use socioeconomic ratings to predict an individual’s actions with a large degree of accuracy. I wish to uncover what really goes into determining a person socioeconomic rating. An Individuals socioeconomic status is very complex. Many factors are involved in determining where the individual falls on the scale. Variables such as wealth, occupation, education, income, privilege, and place of residence are the most heavily weighed.

This report will dig deeper into the metrics behind socioeconomic status. Is it determined by what the individual does or it is something predetermined by birth? Could race be a determining factor in an individual’s socioeconomic status or is it up to individual actions? Are the factors for socioeconomic status the same for every race or are some variables more heavily weighed?

**1. Introduction**

This paper research paper uses the Global Social Survey dataset from 2014. The survey is extremely comprehensive covering topics such as: income, education, children, race, political views. For this survey, 2538 individuals responded.

**2. Data**

The Global Social Survey is given in stata (.dta) format so there were no issues with using the data. The data is cross-sectional. However - since this data is from a survey- some respondents did not reply to every question or did not have an answer. I have tried to account for this by including only variables that I believed described the variables used to determine socioeconomic status while keeping observations as high as possible.

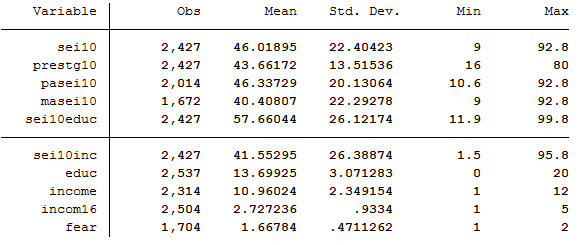


Figure : Summary statistics for primary regression

Socioeconomic status is also not a definitive science. To define socioeconomic status there are many variables used to describe intangible aspects of an individual such as inherited prestige, prejudice, and defunct neighborhoods. This is where the issue with this survey data set lies. For many individuals, some of the variables used to describe intangible aspects do not apply to them and for the survey gave answers such as “not applicable” or “don’t know” which leads to missing variables. This limits the number of observations per regression. To fix this I went through the entire data set finding variables that closely described common aspects of socioeconomics with higher response rates. Due to this, I was able to keep observations up as well as keep the racial proportion close to real.

The variable “fear” is an attempt to control for rough neighborhoods. Explicitly “fear” is whether the respondent feels safe walking in their neighborhood at night. This could be a source for bias. As locals or people who have lived in a certain area for a long period of time feel more safe than perhaps the crime data would show. To help compensate for this bias I included the variable incom16. Incom16 is the household income when the respondent was 16 years old. This will help eliminate some bias and could help measure how good or bad the living conditions were in the respondent’s childhood.

The following graphs display the correlations between some of the largest expected explanatory variables and socioeconomic status:

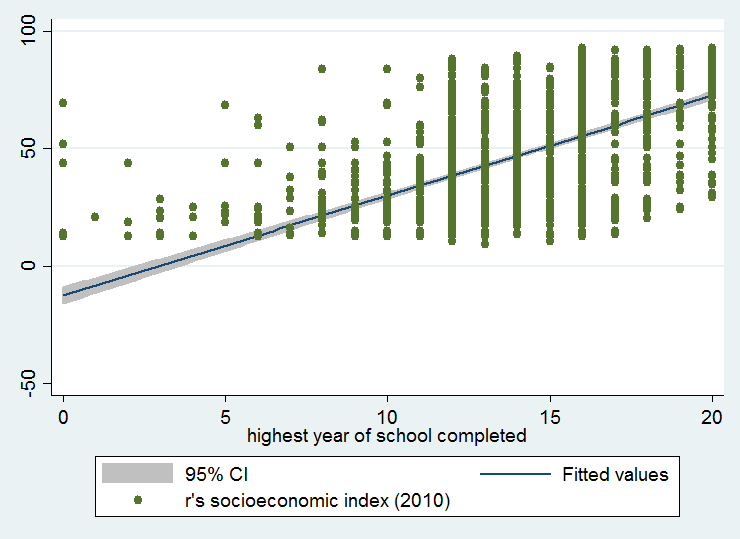


Figure education vs sei

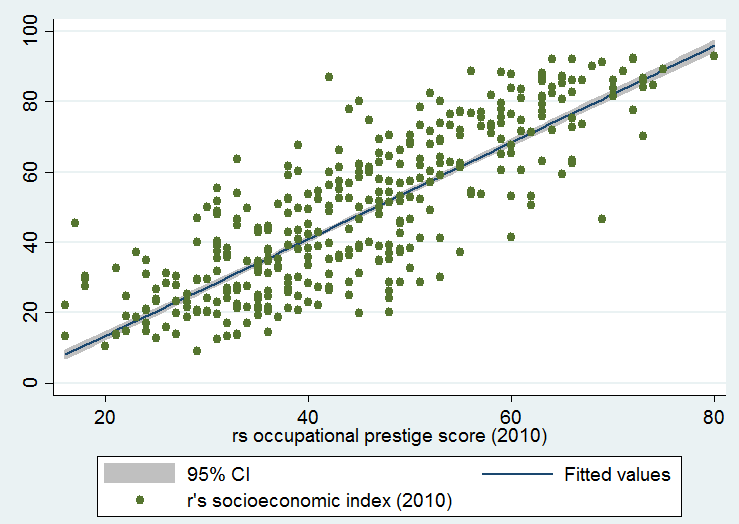
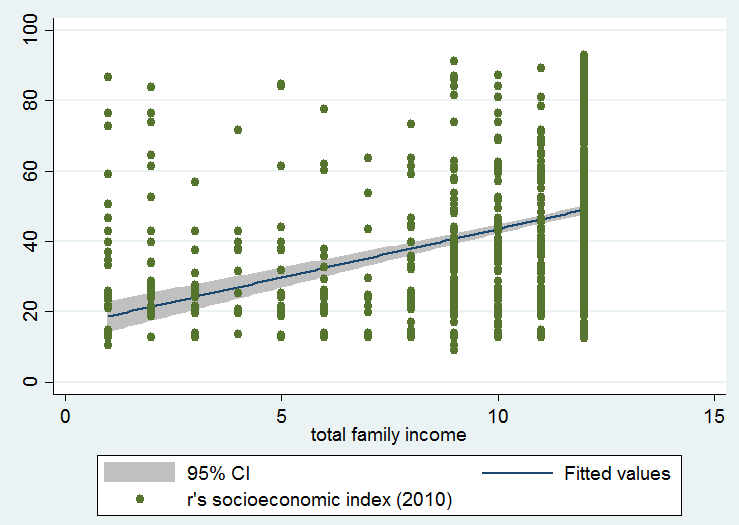


Figure income vs sei Figure job prestige vs sei

The American Psychological Association defines socioeconomic status as: “Socioeconomic status is commonly conceptualized as the social standing or class of an individual or group. It is often measured as a combination of education, income and occupation. Examinations of socioeconomic status often reveal inequities in access to resources, plus issues related to privilege, power and control.” Here is how I controlled for these variables.

|  |  |  |  |
| --- | --- | --- | --- |
| **Income/Wealth** | **Education** | **Occupation** | **Privilege** |
| income | educ | prestg10 | pasei10 |
| sei10inc |  | sei10inc | masei10 |
|  |  | sei10educ | incom16 |
|  |  |  | fear |

I decided to include more variables to explain privilege since it is more vague and immeasurable compared to the others.

**3. Analysis**

Regression/s:

*sei10i = β0 + β1prestg10i + β2pasei10i + β3masei10i + β4sei10educi + β5sei10inci + β6educi + β7incomei + β8income16i + β9feari*  
Then controlling for race:

*sei10i = β0 + β1prestg10i + β2pasei10i + β3masei10i + β4sei10educi + β5sei10inci + β6educi + β7incomei + β8income16i + β9feari if race == 1,2,3 (where 1 =white, 2 = black, 3 = other)*

I first ran the primary regression, then I ran the regression controlling for each race option.

**Table: 1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **Sei10** | **Sei10 (1)** | **Sei10 (2)** | **Sei10 (3)** |
|  |  |  |  |  |
| **prestg10** | 0.07503\*\*\* | 0.08669\*\*\* | 0.016238 | 0.069376\*\* |
|  | (0.011406) | (0.013196) | (0.032958) | (0.03434) |
| **pasei10** | 0.001079 | -0.005943 | 0.005605 | 0.02548\* |
|  | (0.00516) | (0.006007) | (0.016237) | (0.014102) |
| **masei10** | 0.0053399 | 0.0038396 | 0.004731 | 0.01345 |
|  | (0.004376) | (0.004978) | (0.015255) | (0.012434) |
| **sei10educ** | 0.4125704\*\*\* | 0.41145\*\*\* | 0.43048\*\*\* | 0.388743\*\*\* |
|  | (0.00652) | (0.007357) | (0.018164) | (0.02322) |
| **sei10inc** | 0.45688\*\*\* | 0.44989\*\*\* | 0.475305\*\*\* | 0.47132\*\*\* |
|  | (0.005506) | (0.006397) | (0.0135) | (0.020589) |
| **educ** | 0.16589\*\*\* | 0.229142\*\*\* | -0.1339215 | 0.083257 |
|  | (0.03892) | (0.0454) | (0.121838) | (0.10051) |
| **income** | -0.0934\*\* | -0.08717 | -0.048355 | -0.06753 |
|  | (0.04547) | (0.063425) | (0.100838) | (0.09363) |
| **incom16** | -0.19045\* | -0.12831 | -0.058467 | -0.406804 |
|  | (0.112314) | (0.135731) | (0.2683373) | (0.34234) |
| **fear** | -0.09578 | -0.159197 | 0.1480268 | -0.78691 |
|  | (0.1919) | (0.2186047) | (0.520293) | (0.71587) |
| **constant** | -0.96883 | -1.73066\* | 2.107336 | 1.61138 |
|  | (0.73899) | (0.9318) | (1.946658) | (2.2159) |
| **Control for Race** | N/A | white | black | other |
| **Observations** | 774 | 608 | 98 | 68 |
| **Percent of Population** |  | 78.553% | 12.66% | 8.786% |
| **U.S. Census Bureau Racial %’s** |  | 77.1% | 13.3% | 9.6% |
| **R-Squared** | 0.9880 | 0.9878 | 0.9856 | 0.9931 |

\*\*\* denotes p<0.01, \*\* denotes p<0.05, \* denotes p<0.1. Green denotes coefficient values higher than population model. Red denotes the lowest coefficient value for each variable.

Some of these numbers don’t make exact sense. Such as income being negative. I ran a Ramsey RESET test to see if my regression was properly specified. The test produced an f-stat of 90.12 and a p-value of 0.00 which points to specification errors in the regression. I then added polynomial variables for educ (educ^2) and income (income^2). This lowered the f-stat to 74.51 but kept the p-value at 0.00. I then added more polynomial variables (all x^2) for pasei10, masei10, sei10educ, and sei10inc. The result of this was a higher f-stat of 308.11 and p-value of 0.00.

Finally, I arrived at a proper specification for this regression. I included prestg102 (prestg^2), prestg103 (prestg^3), educ2 (educ^2), educ3 (educ^3), income2 (income^2), income3 (income^3), incom162 (incom16^2), inheritedsei (masei10\*pasei10), and workplacesei (sei10educ\*sei10inc). Using these variables gave me an f-stat of 1.26 and a p-value of 0.2854, showing I do not have evidence that the regression isn’t specified correctly.

**Specified Regression:**

*sei10i = β0 + β1prestg10i + β2prestg102i + β3prestg103i + β4pasei10i + β5masei10i + β6sei10educi + β7sei10inci + β8educi + β9educ2i + β10educ3i + β11incomei + β12income2i + β13income3i + β14incom16i + β15incom162i + β16feari+ β17inheritseii + β18workplaceseii if race == 1,2,3 (where 1 =white, 2 = black, 3 = other)*

Running a White test on this new regression gives an f-stat of 35.16 and a p-value of 0.00. This means we have sufficient evidence to reject the null hypothesis of homoscedasticity. Since the regression is found out to be heteroscedastic, robust standard errors will be used. Robust is used to relax the assumption that the errors are distributed identically.

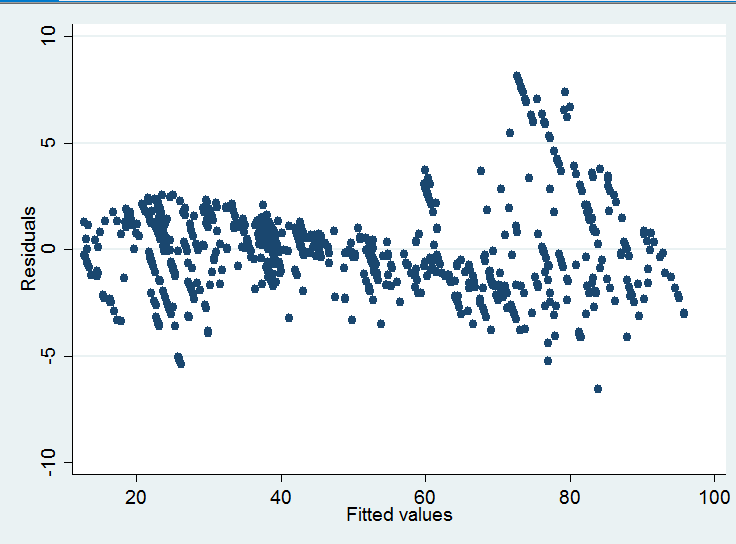


Figure Fitted values vs residuals of the pooled model shows that there is inconsistent variance

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **Sei10** | **Sei10 (1)** | **Sei10 (2)** | **Sei10 (3)** |
| **prestg10** | -0.16993 | -0.114022 | 0.515512 | 0.025689 |
|  | (0.1441) | (0.17944) | (0.513) | (0.26104) |
| **prestg102** | 0.002492 | 0.001579 | -0.020335 | 0.0002709 |
|  | (0.00321) | (0.00398) | (0.01295) | (0.005988) |
| **prestg103** | -8.11e-07 | 4.39e-06 | 0.0002176\*\* | -2.36e-06 |
|  | (0.0000223) | (0.0000276) | (0.0001069) | (0.0000416) |
| **pasei10** | -0.01839\*\* | -0.0182347\* | -0.068872\*\*\* | -0.0298 |
|  | (0.0087397) | (0.0107) | (0.02489) | (0.01877) |
| **masei10** | -0.013403 | -0.00827 | -0.074508\*\*\* | -0.03785 |
|  | (0.0091) | (0.01128) | (0.025785) | (0.023815) |
| **sei10educ** | 0.3645099\*\*\* | 0.35873\*\*\* | 0.381389\*\*\* | 0.34997\*\*\* |
|  | (0.0069) | (0.00789) | (0.020724) | (0.023589) |
| **sei10inc** | 0.3761\*\*\* | 0.362287\*\*\* | 0.422286\*\*\* | 0.373244\*\*\* |
|  | (0.010044) | (0.012363) | (0.020724) | (0.038291) |
| **educ** | 0.839699\*\*\* | 1.08147\* | 0.813471 | 0.93079\* |
|  | (0.29834) | (0.55369) | (1.449915) | (0.49177) |
| **educ2** | -0.092799\*\*\* | -0.11107\*\* | -0.05619 | -0.110815\* |
|  | (0.03138) | (0.05012) | (0.129572) | (0.05603) |
| **educ3** | 0.0030476\*\*\* | 0.003539\*\* | 0.0009265 | 0.00379\*\* |
|  | (0.000968) | (0.00142) | (0.003554) | (0.00178) |
| **income** | 1.68722\*\*\* | 1.21178 | 1.47478 | 3.800027\*\*\* |
|  | (0.504583) | (0.70185) | (1.11599) | (1.227716) |
| **income2** | -0.255234\*\*\* | -0.1765 | -0.22286 | -0.65184\*\*\* |
|  | (0.082244) | (0.1128) | (0.18572) | (0.223795) |
| **income3** | 0.011015\*\*\* | 0.00743 | 0.0096 | 0.030299\*\*\* |
|  | (0.0038177) | (0.005185) | (0.008622) | (0.011143) |
| **incom16** | 0.198287 | 0.29443 | -0.0181933 | 2.0716 |
|  | (0.43547) | (0.50595) | (1.14904) | (1.47178) |
| **incom162** | -0.071104 | -0.08932 | 0.04277 | -0.562607\*\* |
|  | (0.075253) | (0.08915) | (0.17459) | (0.332519) |
| **fear** | -0.125766 | -0.12876 | 0.06807 | -0.7179774 |
|  | (0.1714) | (0.1948) | (0.45845) | (0.671813) |
| **inheritsei** | 0.0004022\*\* | 0.0003032 | 0.001779\*\*\* | 0.00979\*\* |
|  | (0.0001739) | (0.0002059) | (0.000566) | (0.000472) |
| **workplacesei** | 0.0012357\*\*\* | 0.001348\*\*\* | 0.0009618\*\* | 0.00141\*\*\* |
|  | (0.000146) | (0.0001749) | (0.0004679) | (0.00046) |
| **constant** | 4.16635 | 2.9984 | -2.50356 | -0.3825 |
|  | (2.3443) | (3.345) | (7.50608) | (4.55316) |
| **Control for Race** | N/A | white | black | other |
| **Observations** | 774 | 608 | 98 | 68 |
| **% of Population** |  | 78.553% | 12.66% | 8.786% |
| **U.S. Census Bureau Racial %’s** |  | 77.1% | 13.3% | 9.6% |
| **R-Squared** | 0.991 | 0.9908 | 0.9914 | 0.9962 |

**Table: 2** \*\*\* denotes p<0.01, \*\* denotes p<0.05, \* denotes p<0.1. Green denotes coefficient values higher than population model. Red denotes the lowest coefficient value for each variable.

**Income/Wealth: (income, income2, income3, sei10inc, workplacesei)**

It seems that it is the most important for “other” races to have any form of income at all. It is very statistically significant and a coefficient of 3.8 it is very economically significant to the individual’s socioeconomic index score. For middle income earners, it seems that whites and blacks are more advantaged than other races despite not being statistically significant. Then for the top earners, other races experience the largest gain in their socioeconomic index, this is also very statistically and economically significant with a coefficient on the income^3 variable being 0.03.

Sei10inc is the measure the of percentage of workers at the respondent’s workplace who make more than $45k a year. Workplacesei is the interaction between sei10inc and sei10educ (the percentage of coworkers who have some college education). This part is extremely interesting and surprised me. Both variables are extremely statistically significant for all races. For blacks, sei10inc is way more economically significant compared to the other races. While for whites and other, workplacesei were more economically significant than both the population model and for blacks.

Analyzing this it appears that for whites the most important aspect of socioeconomic index scoring is work associations with more educated people and earning middle income. For blacks, how much money the individual’s coworkers make hugely affects their socioeconomic index with the largest coefficient of 0.422, but the education level of their coworkers matters much less than for the other races. For “other” races, having a base income is a huge bump in their index score but unlike the other races they receive a large increase also from being in the top earners. For “other” races, similar to whites and blacks, sei10inc and workplacesei are very statistically significant. “Other” races benefit more from workplacesei than just sei10inc, meaning their index score is related closely to how educated their coworkers are.

**Education: (educ, educ2, educ3)**

Having any formal education for whites and “other” puts them ahead of both the population and blacks, this is both economically and statistically significant. While blacks benefit the least from initial education but also benefit the most from intermediate education, however, none of the coefficients for educ, educ2, and educ3 are statistically significant for blacks. Higher education, such as finishing college and post graduate degrees, is very important for whites and “other” for their socioeconomic index score (statistically and economically significant). It also seems blacks benefit the least in their index score for very high levels of education.

**Occupation: (prestg10, prestg102, prest103, sei10inc, sei10educ, workplacesei)**

For prestige of job the only statistically and truly economically significant variable was prestg103 for blacks. For blacks with a very high prestige job their socioeconomic index score increases much more than other races. Sei10inc and sei10educ are also much more important for blacks than for the other races.

Interestingly for the interaction term workplacesei, blacks have the lowest return. While both whites and “other” have higher than population at statistically significant levels. For blacks either being surround by higher earners or very educated people help their score the most, while a combination of the two is less helpful to their score than for whites and “other”.

**Privilege: (pasei10, masei10, incom16, incom162, fear, inheritsei)**

Privilege is a relatively new word for describing inherited advantages. Sociologists argue that it has a large impact on your socioeconomic score and can even lock you into a certain socioeconomic category. My findings on privilege tend to agree. I found that the effect of inheritance is very different between the races.

Have a single parent with a high socioeconomic score was more detrimental to blacks than to whites and “other”. For whites, having a father with a high index score was both statistically and economically significant but a mother with a high score was not statistically significant. Blacks had negative effects at a very statistically significant level and economically when either masei10 or pasei10 were high individually. Combined, however, using the interaction term inheritsei, blacks benefited the most when both parents are present and had high scores (note: for a single parent inheritsei=0). Whites benefited the least from both parents having high scores. “Other” races benefitted similarly to how blacks did when both parents were present and had high scores.

Fear was a variable I believed would help explain for growing up in bad neighborhoods. In the regression though it was never statistically significant. It is of note though that blacks may see a socioeconomic index benefit from living in a bad neighborhood. While both whites and “other” lose points. I don’t have an explanation for why blacks would benefit from this though. Perhaps they are afforded more scholarship opportunities or access to aid due to affirmative action and the quality of their neighborhood.

For whites and “other” races having a baseline home income as a 16-year-old was economically significant for the socioeconomic score but not statistically significant. For blacks, a basic home income at 16 was a detriment to their index score but once again not statistically significant. Blacks also seemed to benefit the most from having a very high wage earning home, but this was also not statistically significant. The most statistically significant metric was that for “other” races in a high wage earning home at age 16 their index score dropped very significantly. This could be due to social and cultural pressures or moving to America from another country.

**4. Conclusion**

It seems that for different races there are different variables and different levels of importance for variables for their socioeconomic score. An easy explanation would be racism or bigotry, but I do not believe that is entirely the reason for the discrepancies. I think much of the differences can be explained by childhood upbringing and culture.

For example, masei10 and pasei10 are detrimental to all races but especially to blacks. This could be explained by the higher single parent rates for blacks. Statistically, 67% of blacks grow up in a single parent household as opposed to 25% of whites. It could also explain why the interaction term inheritsei is higher for blacks than for other races (it would equal zero if one of the parents were missing). All races have their socioeconomic score boosted by growing up in a home with both a mother and father. Blacks seem to benefit the most from inheritsei due to the larger number of single parent homes.

**Term/variable Definitions:**

sei10 – survey respondents socioeconomic index score based on the 2010 American Community Survey

prestg10 – respondent’s occupational prestige score

pasei10 - survey respondent’s father’s socioeconomic index score based on the 2010 American Community Survey

masei10 - survey respondent’s mother’s socioeconomic index score based on the 2010 American Community Survey

sei10educ – percentage of some college education at respondent’s occupation

sei10inc – percentage of $45k+ earners at respondent’s occupation

educ – respondent’s years of education

income – respondent’s income bracket

incom16 – income of respondent’s home at age 16

fear – Is the respondent afraid to walk around their neighborhood at night

**Explanation for prestige score**: “In 1989 General Social Survey replicated and extended the earlier Hodge-Siegel-Rossi ratings of occupational prestige. The prestige module was designed by the late Robert W. Hodge, Judith Treas and Keiko Nakao in consultation with Smith and Davis. Hodge died in February, 1989 during data collection. Analysis of the data and construction of the NORC/GSS Occupational Prestige Scale was carried out by Nakao and Treas. Their work was supported by NSF. The GSS Board of Overseers Sub-Committee on Occupational Prestige provided advice and guidance to both sets of PIs. Committee members included Richard Campbell, Robert Hauser, Patricia Roos, Paul Siegel, Joe Spaeth, and Andrea Tyree. The occupational prestige module and the NORC/GSS Occupational Prestige Scale are discussed in detail in GSS Methodological Reports No. 69 and 70.3 The NORC/GSS Occupational Prestige scores for the 1980 Census Occupational Classification are given in Appendix F. A machine-readable file that lists prestige score, occupational code, and title of occupational categories is available from ICPSR. A raw data file and SPSS-X control cards that contain the raw prestige ratings that respondents assigned to occupational titles (and not the summary prestige scores for the 1980 Census Occupational Classification) is also available from ICPSR. This file is called "Ratings of Occupational Prestige on the 1989 General Social Survey."

**Explanation for socioeconomic index**: “SEI scores were originally calculated by Otis Dudley Duncan based on NORC's 1947 North-Hatt prestige study and the 1950 U.S. Census. Duncan regressed prestige scores for 45 occupational titles on education and income to produce weights that would predict prestige. This algorithm was then used to calculate SEI scores for all occupational categories employed in the 1950 Census classification of occupations. Similar procedures have been used to produce SEI scores based on later NORC prestige studies and censuses. The GSS contains two sets of SEI scores. Both used procedures similar to those employed by Duncan. For cases coded according to the 1970 US Census codes there are SEI scores developed by Lloyd V. Temme (See Appendix G). These exist for respondent (DOTPRES), spouse (SPDOTPRE), and father (PADOTPRE). For cases coded according to the 1980 US Census codes there are SEI scores developed by Nakao and Treas as part of the GSS's 1989 occupational prestige study (see above). These exist for respondent (SEI), respondent’s first occupation (FIRSTSEI), father (PASEI), mother (MASEI), and spouse (SPSEI). These scores are discussed in GSS Methodological Report No. 74.4”

source: <http://gss.norc.org/documents/codebook/GSS_Codebook_AppendixG.pdf>

Works Cited

Data retrieved from: <http://gss.norc.org/documents/stata/2014_stata.zip>  
  
Census data retrieved from: <https://www.census.gov/quickfacts/>

Single parenthood statistics retrieved from: <http://www.politifact.com/truth-o-meter/statements/2013/jul/29/don-lemon/cnns-don-lemon-says-more-72-percent-african-americ/>

American Psychological Association’s definition of socioeconomics: <http://www.apa.org/topics/socioeconomic-status/>