

University of St.Gallen

School of Management, Economics, Law, Social Sciences and International Affairs

Data Analytics II: PC5

University of St. Gallen

Jonas Husmann | 16-610-917

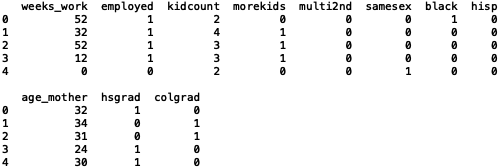
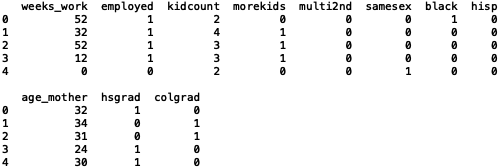
Niklas Leander Kampe | 16-611-618

Prof. Dr. Michael Lechner

March 28, 2022

**Part 1: Descriptive Statistics**

1. Load Data Set

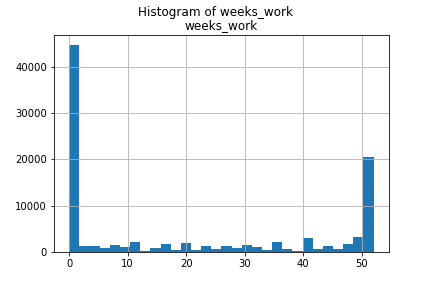
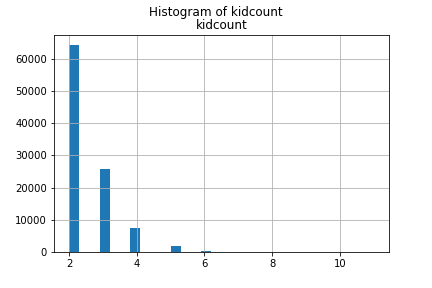


1. Summary Statistics + Plausibility and Patterns

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

According to the summary statistics, the data set does not contain any missing values, which can be identified from the column “na” showing zeros for all variables. Furthermore, the dummy variables “employed”, “morekids”, “multi2nd”, “samesex”, “black”, “hisp”, “hsgrad” and “colgrad” seem to be coded correctly as they show min-values of 0, max-values of 1 and number of unique values of 2. In addition, all values of the “continuous” variables seem to be plausible compared to their intrinsic definitions/descriptions. While looking at the race of the study participants, it seems to be quite homogenous with 12% being Afro-American (“black”) and 3% being Hispanic (“hisp”). In addition, the sample is quite balanced regarding the education, as 48% have high school degrees and 30% having a college degree.



According to the histograms, the variable “kidcount” has the highest mass of observations at the value of 2, which is strictly decreasing up the value of 11. Looking at the distribution, one could think about defining a threshold value of outlying observations to prevent their influence on further econometric analysis. A reasonable threshold could be defined for more than six kids, but requires additional analysis to not omit too much observations. In addition, according to the histogram of “weeks\_work”, the distribution is strongly skewed to the min- and max-extremes. This shows that most of the sample either does not work or follows a full-time job, while the rest of the sample is part-time employed with working times between 0 and 50.

1. Means and Numbers of Observations

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

The table above shows the mean value of the dummy variable “employed” for every number of kids, as well as the number of observations falling into the unique “kidcount” values. First, in accordance with the distribution/histogram of “kidcount”, the number of observations is strictly decreasing in the number of children. Furthermore, the ratio/percentage of being employed, which is defined as the mean value of a dummy variable, is strictly increasing in the number of children until the number of children reaches a value of 9 or more. As the number of observations are fairly small for 9 or more children, and hence the influence on a single observation on the mean is rather large, one could think of applying a threshold on outlying observations on “kidcount” at 9 and more children, as this would underline a strict decrease of employment in the number of children for the whole data set, which would be in accordance with general assumptions that employment becomes less as more children a woman has. Nevertheless, the observations for 8 or less children is expected with participants who have 2 children being most likely employed at a rate of about 60%.

1. Cross Table

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

The table above shows the number of observations which fullfill the unique value combinations for the variables “multi2nd” and “morekids”. The results are indeed plausible, as, first of all, having less or equal than 2 kids (“morekids” = 0) and having the 2nd birth with two or more children at a time (“multi2nd” = 0), which would conclude a total of more than two kids, has zero observations. This underlines full plausibility. Furthermore, most observations (,64’373) lie at the intersection of having not more than two kids and not having a second birth with more than one kids, which is also in accordance with the distribution of “kidcount”. The second highest likelihood (,34’726) lies at the intersection of having more than two kids but not having a second birth with more than one child. Lastly, a minority is observed with having more than two kids and having a second birth with more than one child (901). In total, the sum of all four unique combinations also equals the number of total observations. Hence, the cross-table observations are expected and seem fully plausible.

**Part 2: Homogeneous Effects**

1. OLS Estimation + Causal Relationship

Table

Description automatically generated

The OLS estimation results show that having more than two children has an estimated effect of -5.82 weeks worked per year.

1. 2SLS Estimator + Causal Effect

Table

Description automatically generated