

University of St.Gallen

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

Data Analytics II: PC2

University of St. Gallen

Jonas Husmann | 16-610-917

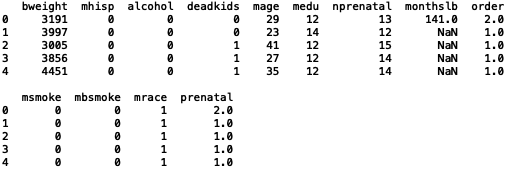
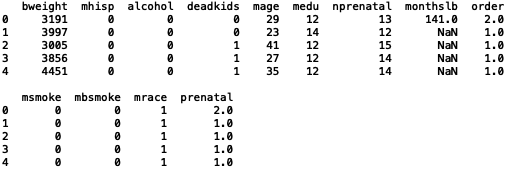
Niklas Leander Kampe | 16-611-618

Prof. Dr. Michael Lechner

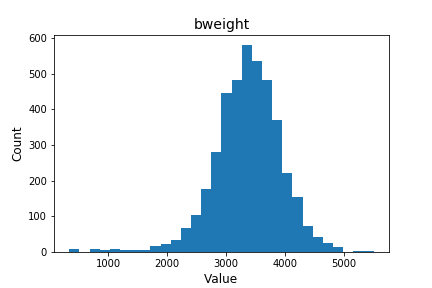
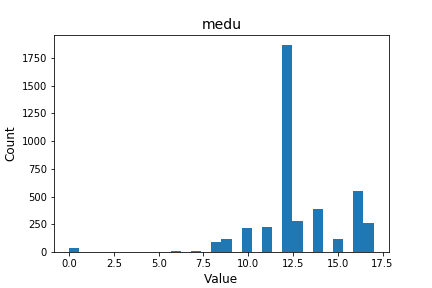
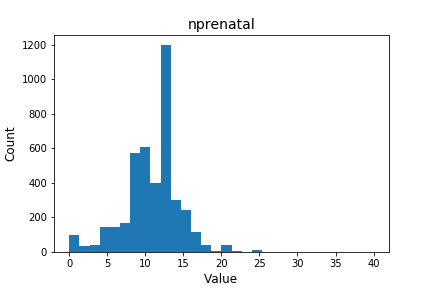
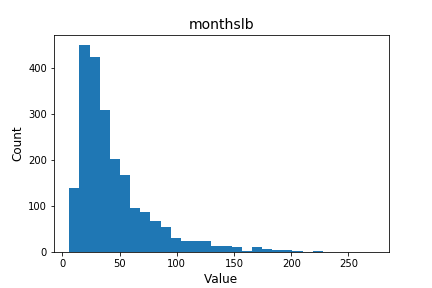
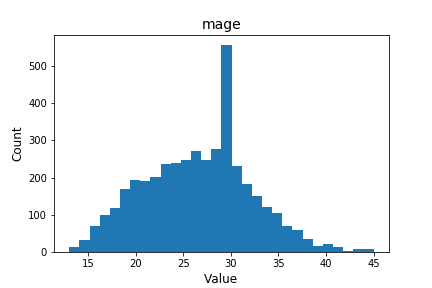
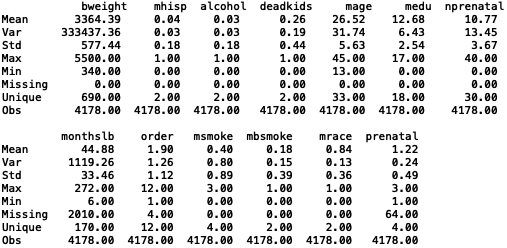
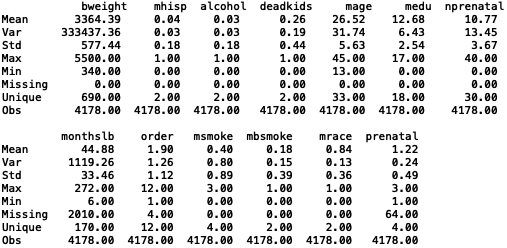
March 07, 2022

**Part 1: Data Preparation**

1. Load Data Set

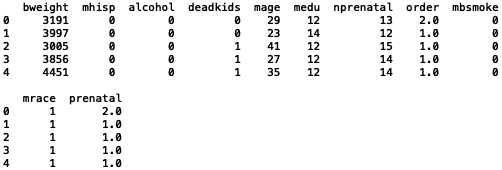
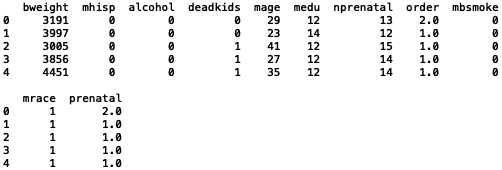


1. Summary Statistics and Histograms + Anomalies or Missing Values



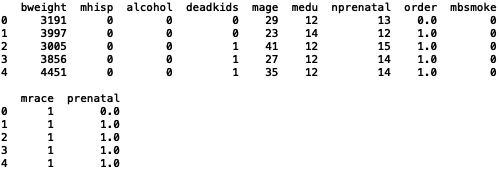
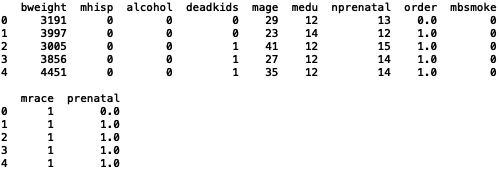
From the summary statistics above, it’s observable that all dummy variables are coded correctly, which can be seen from the min-value of 0, max-value of 1 and the number of unique values of 2. Furthermore, the other variables “bweights”, “mage”, “medu”, “nprenatal”, “monthslb”, “order”, “msmoke” and “prenatal” seem also to have no anomalies since the min-, max- and number of unique values match the variable definition and their description (e.g., variables coded on a 3-step interval have indeed 3 unique values or reasonable birth weights).

1. Variable and Row Drops



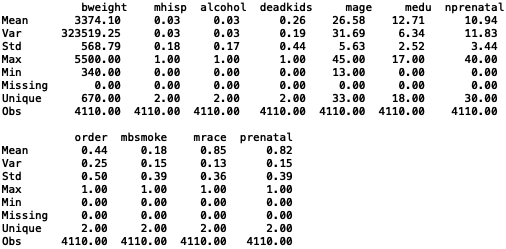
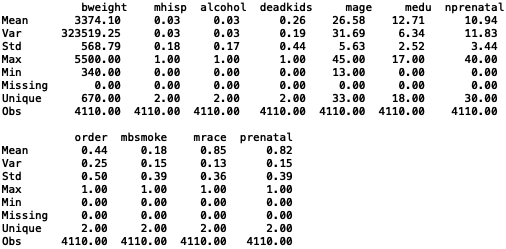
﻿Lorem Ipsum

1. Dummy Variable Recoding



Lorem Ipsum

1. Summary Statistics + Suitability for Economic Analysis



Lorem Ipsum

1. Balancing Checks

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Lorem Ipsum



Lorem Ipsum

**Part 2: ATE Estimation**

1. OLS Function for ATE Estimation without Covariates

A screenshot of a computer

Description automatically generated with low confidence

In the table above we can see the outcome of the OLS without covariates. The estimated ATE is -263.4665 grams, indication that smoking reduces the birthweight by this much on average. This estimated ATE is the same (except for the decimals that are slightly different) as the mean difference that we saw in the balance check in 1f). The underlying identifying assumptions are:

* **Conditional Independence Assumption**: given no covariates included, we assume that no control variable influences both the dependent and independent variable. This is unlikely here and we should include other variables such as *alcohol* (see 1f)
* **Common Support Assumption**: we can consider this to be fine, as we assume no confounding variables
* **Exogeneity of confounders**: assuming no confounding variables, this can be considered fine too
* **Stable Unit Treatment Value Assumption (SUTVA)**: Given that there are different intensities of smoking (0-3 scale) of smoking this is violated

Given that not all the above mentioned assumptions hold, we cannot be sure that the estimated effect is correct. To design an experiment that would remove the confounding it would be necessary to correct for the differences between the treatment and control group, which would allow to remove the selection bias.

1. OLS Function for ATE Estimation with Covariates

As already touched upon in question 2a) we now correct for the differences between the treatment and control group and include the covariates that had a significant and large (above the 10 treshold defined in 1f) difference based on the balance check results. Therefore, we excluded *mhisp* and *mrace* and this OLS regression. We find that the estimated effect of smoking on birthweight is at -226.6595 grams less negative than it was before (see table below).

A screenshot of a computer

Description automatically generated with low confidence

With regards to the assumptions, we again require the same assumptions to hold as in 2a) that is:

* **Conditional Independence Assumption**: Based on the results from 1f) and the included covariates here we assume that no control variable influences both the dependent and independent variable. There are potential other control variables that are unobserved that potentially jointly influence the outcome and selection (will be revisited below)
* **Common Support Assumption**: we can consider this to be fine, as can be seen in the balance check of 1f)
* **Exogeneity of confounders**: This can be considered valid. However, keeping in mind that smoking could potentially also have an influence on other behaviors (such as drinking alcohol).
* **Stable Unit Treatment Value Assumption (SUTVA)**: Same as in 2a) this is violated given that there are different intensities of smoking (0-3 scale) of smoking.

Further unobservable covariates that could threaten our identification strategy could be:

* The educational level of the father
* The household income level
* Mental health of the mother
* Parent’s birthweight