Notes from 2020-04-21. Conference call Georg, Marianne, Mark

**NHANES project**

Intended statistical analysis plan

Outcome: MVPA (minutes of moderate/vigorous physical activity)

Statistical model(s):

1. Linear regression with explanatory variables age, gender, education, smoking, alcohol consumption, BMI, comorbidities (cancer, CHF, stroke), cholesterol (total, HDL). Study partial R-squared.
2. Linear regression model to study the role of systolic blood pressure in addition to the variables above.

Purpose: IDA for exploratory regression model

IDA plan

1. Univariate distributions
2. Multivariate distributions
3. Logistic regression models with comorbidities as outcomes and age as explanatory variable to study loess curve, stratified by gender.
4. Total cholesterol vs HDL
5. Missing values (e.g. blood pressure)

**Presentation for ISCB** will very likely be virtual.

Power point or Beamer or (a favorite template?)

* Introduce the two data sets and analysis plan
* “Worst case scenario” IDA is performed with non-transparent impact on analysis.
* What is IDA for regression (in the figure on IDA steps, we focus on data screening and updating analysis plan, possibly mention reporting)
* How do we interpret an explanatory variable (e.g. age in the NHANES data) correctly?
* Examples of reporting

Thoughts for collaborative presentation to make it more interesting for the audience? For example, with specifically placed comments or questions?