

My Reproducible Script

Heleen Brüggem

Introduction

This document reproduces the findings from the article by Boulesteix, Groenwold, Abrahamowicz et al. (2020). The data used for this article is the 2015–2016 National Health and Nutrition Examination Survey (NHANES) and can be downloaded from [NHANES](#).

First, the data was subsetting and merged into one data set. The following code does there mentioned steps.

```
# Libraries
library(Hmisc)
library(mice)
library(tidyverse)

# Setting the working directory
setwd('/Users/Heleen/Desktop/Universiteit Utrecht/Methodology & Statistics for the Behaviour

# Loading the data
d1 <- Hmisc::sasxport.get("Data/DEMO_I.XPT")
d2 <- Hmisc::sasxport.get("Data/BPX_I.XPT")
d3 <- Hmisc::sasxport.get("Data/BMX_I.XPT")
d4 <- Hmisc::sasxport.get("Data/GHB_I.XPT")
d5 <- Hmisc::sasxport.get("Data/TCHOL_I.XPT")

# Subsetting data
d1.t <- subset(d1,select=c("seqn","riagendr","ridageyr"))
d2.t <- subset(d2,select=c("seqn","bpxsy1"))
d3.t <- subset(d3,select=c("seqn","bmxbmi"))
d4.t <- subset(d4,select=c("seqn","lxbgh"))
d5.t <- subset(d5,select=c("seqn","lbdtsi"))

# Merging data sets
```

```
d <- merge(d1.t,d2.t)
d <- merge(d,d3.t)
d <- merge(d,d4.t)
d <- merge(d,d5.t)
```

Renaming variables

For readability and ease, the following code renames the variables `ridageyr`, `riagendr`, `bpxsy1`, `bmxbmi`, `lbxgh` and `lbdtsi` to `age`, `sex`, `bp`, `bmi`, `HbA1C` and `chol` respectively. Respondents below 18 were coded as missing.

```
d$age <- d$ridageyr # ridageyr --> age
d$sex <- d$riagendr # riagendr --> sex
d$bp <- d$bpxsy1 # bpxsy1 --> bp
d$bmi <- d$bmxbmi # bmxbmi --> bmi
d$HbA1C <- d$lbxgh # lbxgh --> HbA1C
d$chol <- d$lbdtsi # lbdtsi --> chol
d$age[d$age<18] <- NA # coding respondents below 18 as missing
```

Selecting complete cases

The data was subsetted to only include complete cases and the variables `age`, `sex`, `bmi`, `HbA1C` and `bp`.

```
dc <- cc(subset(d,select=c("age","sex","bmi","HbA1C","bp")))
```

Analyses

The first analysis was a linear regression with blood pressure (`bp`) as dependent variable and `HbA1C`, `age` and `sex` as independent variables. The second linear regression with blood pressure as dependent variable and `HbA1C`, `bmi`, `age` and `sex` as independent variables. As can be seen from the next code block

```
summary(lm(bp ~ HbA1C + age + as.factor(sex), data=dc))
confint(lm(bp ~ HbA1C + age + as.factor(sex), data=dc))
summary(lm(bp ~ HbA1C + bmi + age + as.factor(sex), data=dc))
confint(lm(bp ~ HbA1C + bmi + age + as.factor(sex), data=dc))
```

Simulation of measurement error

Next, a simulation study was performed to investigate to possible measurement error associated with the linear models. First, the original recordings of the data were assumed to be measured without error.

```
ref <- lm(bp ~ HbA1C + bmi + age + as.factor(sex), data=dc)$coef[2]
n.sim <- 1e3
perc.me.exp <- seq(0,.5,.1)
perc.me.conf<- seq(0,.5,.1)
scenarios <- expand.grid(perc.me.exp,perc.me.conf)
var.exp <- var(dc$HbA1C)
var.conf <- var(dc$bmi)
n <- dim(dc)[1]
beta.hat <- matrix(ncol=dim(scenarios)[1], nrow=n.sim)
for (k in 1:n.sim){
  print(k)
  set.seed(k)
  for (i in 1:dim(scenarios)[1]){
    var.me.exp <- var.exp*scenarios[i,1]/(1-scenarios[i,1])
    var.me.conf <- var.conf*scenarios[i,2]/(1-scenarios[i,2])
    dc$HbA1C.me <- dc$HbA1C + rnorm(dim(dc)[1], 0, sqrt(var.me.exp) )
    dc$bmi.me <- dc$bmi + rnorm(dim(dc)[1], 0, sqrt(var.me.conf) )
    beta.hat[k,i] <- lm(bp ~ HbA1C.me + age + bmi.me + as.factor(sex), data=dc)$coef[2]
  }}
}}
```

Results

The following code prints the results of the simulation study.

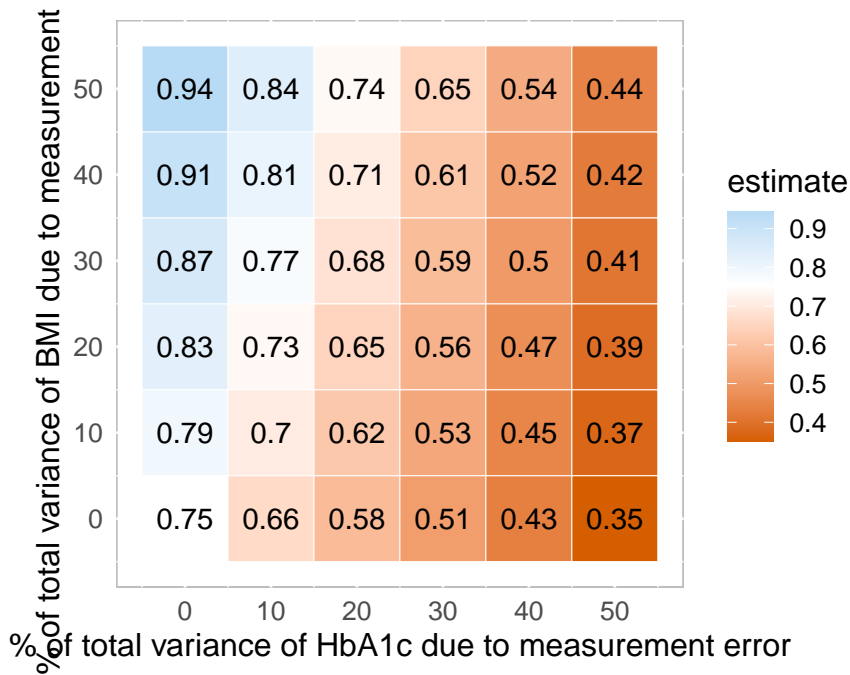
```
tot.mat <- cbind(100*scenarios,apply(beta.hat,2,mean))
colnames(tot.mat) <- c("me.exp","me.conf","estimate")
FIGURE <- ggplot(tot.mat, aes(me.exp, me.conf)) +
  geom_tile(color="white",aes(fill = estimate)) +
  geom_text(aes(label = round(estimate, 2))) +
  scale_fill_gradient2(low="#D55E00",mid="white",high = "#56B4E9", midpoint=ref) +
  labs(x=paste("% of total variance of HbA1c due to measurement error"),
       y=paste("% of total variance of BMI due to measurement error")) +
  coord_equal()+
  scale_y_continuous(breaks=unique(tot.mat[,1]))+
  scale_x_continuous(breaks=unique(tot.mat[,1]))+
  theme(panel.background = element_rect(fill='white', colour='grey'),
```

```

plot.title=element_text(hjust=0),
axis.ticks=element_blank(),
axis.title=element_text(size=12),
axis.text=element_text(size=10),
legend.title=element_text(size=12),
legend.text=element_text(size=10))

```

FIGURE



```

# savePlot("Figure_STRATOS.tif", type="tif")

```

References

Boulesteix, Anne-Laure, Rolf Hh Groenwold, Michal Abrahamowicz, Harald Binder, Matthias Briel, Roman Hornung, Tim P Morris, Jörg Rahnenführer, and Willi Sauerbrei. 2020. "Introduction to Statistical Simulations in Health Research." *BMJ Open* 10 (12): e039921. <https://doi.org/10.1136/bmjopen-2020-039921>.