

# My Reproducible Manuscript: Exercise 2

Martijn L.G. van Dam

## Intro

This documents aims at reproducing the results of a simulation study by Boulesteix et al. (2020), provided by Hanne and Gerko for the Markup languages course.

## Load Libraries:

```
# Packages
packages <- c("mice", "Hmisc", "tidyverse")

# Install above-mentioned packages if not yet installed
installed_packages <- packages %in% rownames(installed.packages())
if (any(installed_packages == FALSE)) {
  install.packages(packages[!installed_packages])
}

# Load the packages
invisible(lapply(packages, library, character.only = TRUE))
```

```
d1 <- sasxport.get("./data/DEMO_I.xpt")
```

Processing SAS dataset DEMO\_I ..

```
d2 <- sasxport.get("./data/BPX_I.xpt")
```

Processing SAS dataset BPX\_I ..

```
d3 <- sasxport.get("./data/BMX_I.xpt")
```

Processing SAS dataset BMX\_I ..

```
d4 <- sasxport.get("./data/GHB_I.xpt")
```

Processing SAS dataset GHB\_I ..

```
d5 <- sasxport.get("./data/TCHOL_I.xpt")
```

Processing SAS dataset TCHOL\_I ..

```
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","lbxgh"))
d5.t <- subset(d5,select=c("seqn","lbdtsi"))
d <- merge(d1.t,d2.t)
d <- merge(d,d3.t)
d <- merge(d,d4.t)
d <- merge(d,d5.t)
# =====
# rename variables:
# RIAGENDR - Gender
# RIDAGEYR - Age in years at screening
# BPXSY1 - Systolic: Blood pres (1st rdg) mm Hg
# BMXBMI - Body Mass Index (kg/m**2)
# LBDTCSI - Total Cholesterol (mmol/L)
# LBXGH - Glycohemoglobin (%)
d$age <- d$ridageyr
d$sex <- d$riagendr
d$bp <- d$bpxsy1
d$bmi <- d$bmxbmi
d$HbA1C <- d$lbxgh
d$chol <- d$lbdtsi
d$age[d$age<18] <- NA
# =====
# select complete cases:
dc <- cc(subset(d,select=c("age","sex","bmi","HbA1C","bp")))
# analysis:
summary(lm(bp ~ HbA1C + age + as.factor(sex), data=dc))
```

```
Call:
lm(formula = bp ~ HbA1C + age + as.factor(sex), data = dc)

Residuals:
Systolic: Blood pres (1st rdg) mm Hg
      Min       1Q   Median       3Q      Max
-49.887 -10.509  -1.378   8.491 107.583

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   98.75149    1.21418  81.332 < 2e-16 ***
HbA1C          1.12638    0.20291   5.551 2.98e-08 ***
age            0.44486    0.01284  34.648 < 2e-16 ***
as.factor(sex)2 -3.24792    0.45164  -7.191 7.34e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.1 on 5088 degrees of freedom
Multiple R-squared:  0.2305,    Adjusted R-squared:  0.23
F-statistic:  508 on 3 and 5088 DF,  p-value: < 2.2e-16
```

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

```
              2.5 %      97.5 %
(Intercept)  96.3711755 101.1317982
HbA1C         0.7285836  1.5241825
age           0.4196932  0.4700355
as.factor(sex)2 -4.1333281 -2.3625106
```

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

```
Call:
lm(formula = bp ~ HbA1C + bmi + age + as.factor(sex), data = dc)

Residuals:
Systolic: Blood pres (1st rdg) mm Hg
      Min       1Q   Median       3Q      Max
-51.068 -10.251  -1.504   8.264 107.410
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	92.65583	1.39320	66.506	< 2e-16 ***
HbA1C	0.75177	0.20596	3.650	0.000265 ***
bmi	0.28632	0.03282	8.724	< 2e-16 ***
age	0.44586	0.01275	34.979	< 2e-16 ***
as.factor(sex)2	-3.63115	0.45049	-8.060	9.4e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.98 on 5087 degrees of freedom

Multiple R-squared: 0.2418, Adjusted R-squared: 0.2412

F-statistic: 405.7 on 4 and 5087 DF, p-value: < 2.2e-16

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

	2.5 %	97.5 %
(Intercept)	89.9245592	95.3871089
HbA1C	0.3479966	1.1555348
bmi	0.2219815	0.3506673
age	0.4208695	0.4708464
as.factor(sex)2	-4.5143014	-2.7479929

```
# =====
```

```
# simulation of measurement error:
```

```
#| output: FALSE
```

```
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]
}}

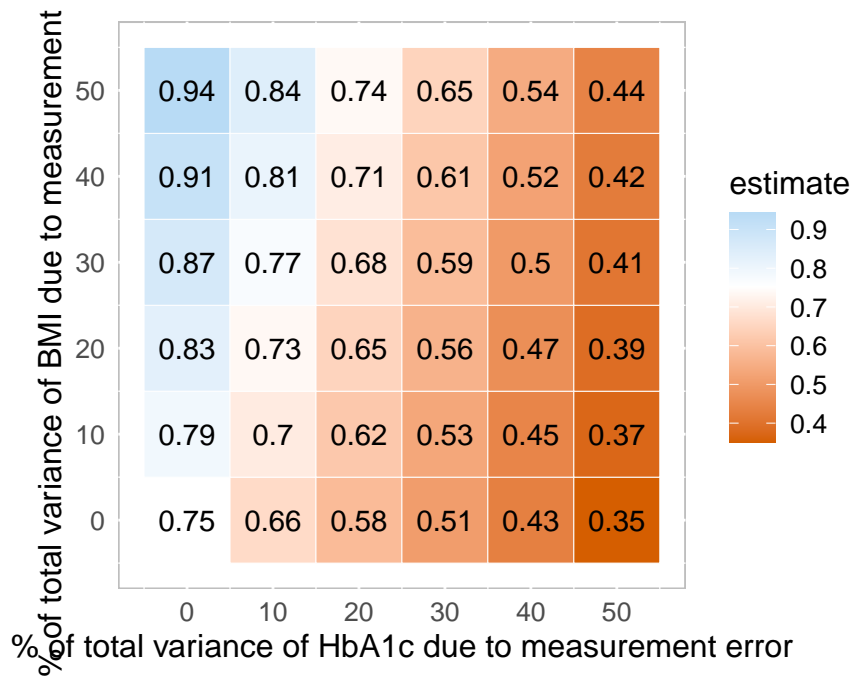
```

```

# =====
# create figure:
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



```
ggsave(filename = "./results/Figure.tif", plot = last_plot(), device = "tiff")
```

Saving 5.5 x 3.5 in image

Boulesteix, Anne-Laure, Rolf H. H. Groenwold, Michal Abrahamowicz, et al. 2020. "Introduction to Statistical Simulations in Health Research." *BMJ Open* 10: e039921. <https://doi.org/10.1136/bmjopen-2020-039921>.