Reproduction of a Small Statistical Simulation
Excercise 2 in week 4

Sophie Hetche

2024-12-28

R CODE

This is a small scale simulation study to investigate impact of measurement error on (continuous) exposure and/or (continuous) confounding variable

Loading the Necessary Libraries

note: uninstalled libraries can be installed by using `install.packages("LibraryName")`

```
# List of required packages
required_packages <- c("Hmisc", "mice", "tidyverse")

# Check if the respective packages are installed, and install if necessary
for (pkg in required_packages) {
  if (!requireNamespace(pkg, quietly = TRUE)) {
    install.packages(pkg)
  }

  # Load the library of the packages
  library(pkg, character.only = TRUE)
}
```

Loading the data

The data can be downloaded in xpt form from this link.

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$chol <- d$lbdtcsi
d$age[d$age<18] <- NA
```

Select Complete Cases

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

Performing the two Regression Analyses

```
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
```

```r
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 the Measurement Error

note: Include `cache=TRUE` in the code chunk header to avoid recomputing when rerunning the document

```r
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]
}}
```

```r
summary(beta.hat)
```

```
      V1                V2                V3                V4
 Min.   :0.7518   Min.   :0.4201   Min.   :0.3026   Min.   :0.09733
 1st Qu.:0.7518   1st Qu.:0.6207   1st Qu.:0.5276   1st Qu.:0.44249
 Median :0.7518   Median :0.6614   Median :0.5846   Median :0.50937
```
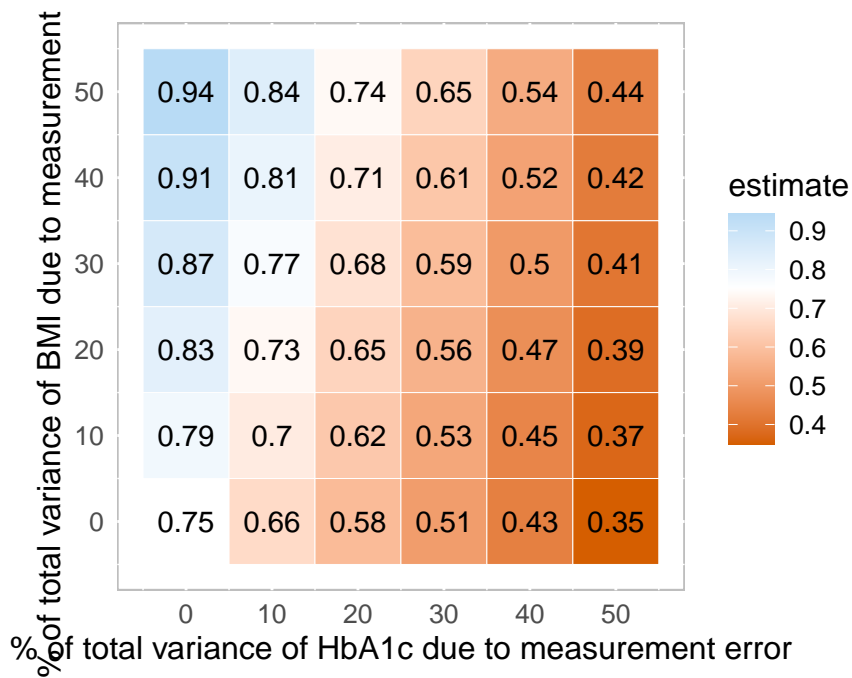
```
Mean   :0.7518   Mean   :0.6628   Mean   :0.5825   Mean   :0.50724
3rd Qu.:0.7518   3rd Qu.:0.7057   3rd Qu.:0.6399   3rd Qu.:0.57376
Max.   :0.7518   Max.   :0.8877   Max.   :0.8707   Max.   :0.84682
      V5               V6               V7               V8
Min.   :0.09439  Min.   :0.007166  Min.   :0.7493   Min.   :0.4846
1st Qu.:0.35746  1st Qu.:0.276282  1st Qu.:0.7808   1st Qu.:0.6599
Median :0.43630  Median :0.355739  Median :0.7906   Median :0.7004
Mean   :0.43384  Mean   :0.350249  Mean   :0.7906   Mean   :0.7035
3rd Qu.:0.50928  3rd Qu.:0.420475  3rd Qu.:0.8003   3rd Qu.:0.7480
Max.   :0.82219  Max.   :0.706103  Max.   :0.8316   Max.   :0.9076
      V9               V10              V11              V12
Min.   :0.3394   Min.   :0.2024   Min.   :0.1651   Min.   :0.02396
1st Qu.:0.5585   1st Qu.:0.4628   1st Qu.:0.3842   1st Qu.:0.30246
Median :0.6156   Median :0.5318   Median :0.4490   Median :0.37497
Mean   :0.6161   Mean   :0.5320   Mean   :0.4479   Mean   :0.37217
3rd Qu.:0.6743   3rd Qu.:0.5977   3rd Qu.:0.5116   3rd Qu.:0.43935
Max.   :0.9335   Max.   :0.8175   Max.   :0.7766   Max.   :0.68417
      V13              V14              V15              V16
Min.   :0.7538   Min.   :0.4698   Min.   :0.3930   Min.   :0.2390
1st Qu.:0.8158   1st Qu.:0.6855   1st Qu.:0.5841   1st Qu.:0.4975
Median :0.8300   Median :0.7343   Median :0.6469   Median :0.5616
Mean   :0.8293   Mean   :0.7338   Mean   :0.6460   Mean   :0.5608
3rd Qu.:0.8437   3rd Qu.:0.7826   3rd Qu.:0.7066   3rd Qu.:0.6271
Max.   :0.8860   Max.   :0.9464   Max.   :0.9289   Max.   :0.8425
      V17              V18              V19              V20
Min.   :0.1434   Min.   :0.05521  Min.   :0.8059   Min.   :0.5131
1st Qu.:0.4028   1st Qu.:0.31293  1st Qu.:0.8520   1st Qu.:0.7262
Median :0.4719   Median :0.38582  Median :0.8682   Median :0.7690
Mean   :0.4712   Mean   :0.38804  Mean   :0.8679   Mean   :0.7706
3rd Qu.:0.5404   3rd Qu.:0.46053  3rd Qu.:0.8834   3rd Qu.:0.8165
Max.   :0.8041   Max.   :0.73510  Max.   :0.9353   Max.   :1.0120
      V21              V22              V23              V24
Min.   :0.3991   Min.   :0.2586   Min.   :0.1295   Min.   :-0.003231
1st Qu.:0.6291   1st Qu.:0.5179   1st Qu.:0.4265   1st Qu.: 0.339181
Median :0.6767   Median :0.5909   Median :0.4982   Median : 0.408294
Mean   :0.6810   Mean   :0.5875   Mean   :0.4990   Mean   : 0.409386
3rd Qu.:0.7387   3rd Qu.:0.6497   3rd Qu.:0.5684   3rd Qu.: 0.482525
Max.   :0.9820   Max.   :0.8806   Max.   :0.8153   Max.   : 0.747951
      V25              V26              V27              V28
Min.   :0.8172   Min.   :0.5853   Min.   :0.4303   Min.   :0.3192
1st Qu.:0.8914   1st Qu.:0.7652   1st Qu.:0.6516   1st Qu.:0.5491
Median :0.9050   Median :0.8083   Median :0.7091   Median :0.6141
Mean   :0.9061   Mean   :0.8088   Mean   :0.7083   Mean   :0.6127
```

```
3rd Qu.:0.9225   3rd Qu.:0.8517   3rd Qu.:0.7651   3rd Qu.:0.6745
Max.   :0.9812   Max.   :1.0372   Max.   :0.9634   Max.   :0.9076
     V29              V30              V31              V32
Min.   :0.1829   Min.   :0.08584   Min.   :0.8528   Min.   :0.5751
1st Qu.:0.4476   1st Qu.:0.35787   1st Qu.:0.9282   1st Qu.:0.7965
Median :0.5242   Median :0.42073   Median :0.9433   Median :0.8425
Mean   :0.5211   Mean   :0.42392   Mean   :0.9432   Mean   :0.8431
3rd Qu.:0.5929   3rd Qu.:0.49451   3rd Qu.:0.9600   3rd Qu.:0.8884
Max.   :0.7760   Max.   :0.78509   Max.   :1.0201   Max.   :1.0373
     V33              V34              V35              V36
Min.   :0.4838   Min.   :0.3089   Min.   :0.2284   Min.   :0.1106
1st Qu.:0.6779   1st Qu.:0.5807   1st Qu.:0.4743   1st Qu.:0.3738
Median :0.7376   Median :0.6488   Median :0.5414   Median :0.4426
Mean   :0.7366   Mean   :0.6458   Mean   :0.5435   Mean   :0.4422
3rd Qu.:0.7952   3rd Qu.:0.7175   3rd Qu.:0.6123   3rd Qu.:0.5098
Max.   :1.0001   Max.   :0.9889   Max.   :0.9302   Max.   :0.7627
```

Create Figure of the Impact of the Measurement Error

```r
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
```

The plot can be saved with `savePlot("Figure_STRATOS.tif", type="tif")`

END OF R CODE