Simulation

Abigail McGrory

2024-11-11

```r
## set simulation parameters
n <- 500     # sample size
pz <- 0.2    # probability of Z = 1
alpha0 <- 0  # logit probability of x = 1 in non-smokers (z = 0)
alpha1 <- 1  # log odds ratio of x = 1 in smokers (z = 1) vs non-smokers
beta0 <- -3  # logit prob of y = 1 in non-coffee drinkers (x = 0) and non-smokers (z = 0)
beta1 <- 0
beta2 <- 2

simulation <- function(n, pz, alpha0, alpha1, beta0, beta1, beta2){

## generate confounder Z from a binomial distribution
z <- rbinom(n, size = 1, prob = pz)
## compute probability of observing X = 1 from the inverse logit function
px <- exp(alpha0 + alpha1 * z) / (1 + exp(alpha0 + alpha1 * z))
## randomly generate binary variable X from the above probability
x <- rbinom(n, size = 1, prob = px)
## randomly generate binary variable Y from the inverse logistic function
py <- exp(beta0 + beta1 * x + beta2 * z) / (1 + exp(beta0 + beta1 * x + beta2 * z))
y <- rbinom(n, size = 1, prob = py)

## combine three random variables into a data frame
dat <- data.frame(lung = y, coffee = x, smoke = z)

dat[1:10,]


## fit unadjusted logistic regression model
unadj.mod <- glm(lung ~ coffee, data = dat, family = "binomial")
unadj.p <- summary(unadj.mod)$coef[2,4]
unadj.rej <- (unadj.p < 0.05)*1

## fit adjusted logistic regression model
adj.mod <- glm(lung ~ coffee + smoke, data = dat, family = "binomial")
adj.p <- summary(adj.mod)$coef[2,4]
adj.rej <- (adj.p < 0.05)*1

return(list(unadj.rej, adj.rej))
}
```

```r
library(foreach)
library(doParallel)
```

```
## Loading required package: iterators
```

```
## Loading required package: parallel
```

```r
# set parameters
n_sim <- 1000
ncores <- detectCores() - 1
cl <- makeCluster(ncores)
registerDoParallel(cl)

Sim_Res <- foreach(k = 1:n_sim, .combine = rbind,
                   .packages = c("boot", "dplyr")
) %dopar% {
  set.seed(k + 123)
  simulation(n = 500 ,    # sample size
pz = 0.2 ,   # probability of Z = 1
alpha0 = 0 , # logit probability of x = 1 in non-smokers (z = 0)
alpha1 = 0 , # log odds ratio of x = 1 in smokers (z = 1) vs non-smokers
beta0 = -3 , # logit prob of y = 1 in non-coffee drinkers (x = 0) and non-smokers (z = 0)
beta1 = 0,
beta2 = 2)
}


results_alpha.0 <- Sim_Res
```

```r
# set parameters
n_sim <- 1000
ncores <- detectCores() - 1
cl <- makeCluster(ncores)
registerDoParallel(cl)

Sim_Res <- foreach(k = 1:n_sim, .combine = rbind,
                   .packages = c("boot", "dplyr")
) %dopar% {
  set.seed(k + 123)
  simulation(n = 500 ,    # sample size
pz = 0.2 ,   # probability of Z = 1
alpha0 = 0 , # logit probability of x = 1 in non-smokers (z = 0)
alpha1 = 1 , # log odds ratio of x = 1 in smokers (z = 1) vs non-smokers
beta0 = -3 , # logit prob of y = 1 in non-coffee drinkers (x = 0) and non-smokers (z = 0)
beta1 = 0,
beta2 = 2)
}


results_alpha.1 <- Sim_Res
```

```r
# set parameters
n_sim <- 1000
ncores <- detectCores() - 1
cl <- makeCluster(ncores)
registerDoParallel(cl)

Sim_Res <- foreach(k = 1:n_sim, .combine = rbind,
                   .packages = c("boot", "dplyr")
) %dopar% {
  set.seed(k + 123)
  simulation(n = 500 ,    # sample size
pz = 0.2 ,   # probability of Z = 1
alpha0 = 0 , # logit probability of x = 1 in non-smokers (z = 0)
alpha1 = 2 , # log odds ratio of x = 1 in smokers (z = 1) vs non-smokers
beta0 = -3 , # logit prob of y = 1 in non-coffee drinkers (x = 0) and non-smokers (z = 0)
beta1 = 0,
beta2 = 2)
}


results_alpha.2 <- Sim_Res
```

```r
library(knitr)
library(kableExtra)
```

```
## Warning: package 'kableExtra' was built under R version 4.4.1
```

```r
df <-data.frame("alpha"=c(0,1,2), "Adjusted" = c(mean(unlist(results_alpha.0[,2])), mean(unlist(results
df %>%
  kable(caption = "Type 1 Error Rates") %>%
  kable_styling(full_width = FALSE)
```

Table 1: Type 1 Error Rates

alpha	Adjusted	Unadjusted
0	0.040	0.044
1	0.037	0.233
2	0.033	0.566