STAT 98 Project 2 Simulation

Chris Canzano, Victoria Li, and Henry Wu

March 17, 2024

```r
library(mgcv) # adds the package for multivariate normal data generation
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.9-0. For overview type 'help("mgcv-package")'.
```

```r
set.seed(98) # set seed for reproducibility
N <- 10000 # number of trials in each configuration - reduce this number to like 100 when you do test r
n <- 100 # sample size
k <- 4 # number of predictor variables
cases <- c('treatment', 'control', 'all') # 'all' refers to the configuration with strong multicollinea
treatment_rhos <- c(0, 0.7, -0.7, 0.99, -0.99) # correlations in configuration with collinearity betwee
control_rhos <- c(0.99, -0.99) # correlations in configuration with collinearity between two controls
var_epsilons <- c(1, 25) # variances of error term - let me know if I should change these to be more re
coef_sets <- list(c(2, 1, 0.5, 0.5), c(2, -1, -0.5, -0.5), c(0.5, 0.5, 1, 2)) # sets of coefficients -

for (case in cases) {
  if (case == 'treatment') {
    rhos <- treatment_rhos
  }
  else if (case == 'control') {
    rhos <- control_rhos
  }
  else if (case == 'all') {
    rhos <- c(1) # only one case which does not have variable rhos
  }

  for (rho in rhos) {
    if (case == 'treatment') {
      Sigma <- matrix(c(1, rho, 0, 0,
                        rho, 1, 0, 0,
                        0, 0, 1, 0,
                        0, 0, 0, 1), 4, 4) # correlation matrix for collinearity between treatment and
    }
    else if (case == 'control') {
      Sigma <- matrix(c(1, 0, 0, 0,
                        0, 1, rho, 0,
                        0, rho, 1, 0,
                        0, 0, 0, 1), 4, 4) # correlation matrix for collinearity between two controls
    }
    else if (case == 'all') {
      Sigma <- matrix(c(1, 0.992, 0.621, 0.465,
                        0.992, 1, 0.604, 0.446,
```

```
                        0.621, 0.604, 1, -0.177,
                        0.465, 0.446, -0.177, 1), 4, 4) # correlation matrix for strong multicollineari
    }
    print('Correlation matrix:')
    print(Sigma)
    # print(paste('Theoretical condition number:', 1/rcond(Sigma)))
    for (coefs in coef_sets) {
      print(paste('Coefficients: [', paste(coefs, collapse=', '), ']', sep=''))
      for (var_epsilon in var_epsilons) {
        print(paste('Variance of error term: =', var_epsilon))
        for (i in 1:N) {
          X <- rmvn(n, rep(0, k), Sigma) # generation of multivariate normal data
          # print(X[1:10,]) # first 10 observations of data
          # print(paste('Condition number:', 1/rcond(t(scale(X)) %*% scale(X)))) # using empirical corr
          epsilon <- rnorm(n, 0, sqrt(var_epsilon)) # generation of error term
          y <- X %*% coefs + epsilon # generation of response variable
          # print(y[1:10]) # first 10 observations of response variable
        }
      }
    }
  }
  print('')
}
```

```
## [1] "Correlation matrix:"
##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    0
## [2,]    0    1    0    0
## [3,]    0    0    1    0
## [4,]    0    0    0    1
## [1] "Coefficients: [2, 1, 0.5, 0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [2, -1, -0.5, -0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [0.5, 0.5, 1, 2]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Correlation matrix:"
##      [,1] [,2] [,3] [,4]
## [1,]  1.0  0.7    0    0
## [2,]  0.7  1.0    0    0
## [3,]  0.0  0.0    1    0
## [4,]  0.0  0.0    0    1
## [1] "Coefficients: [2, 1, 0.5, 0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [2, -1, -0.5, -0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [0.5, 0.5, 1, 2]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
```

```
## [1] "Correlation matrix:"
##      [,1] [,2] [,3] [,4]
## [1,]  1.0 -0.7    0    0
## [2,] -0.7  1.0    0    0
## [3,]  0.0  0.0    1    0
## [4,]  0.0  0.0    0    1
## [1] "Coefficients: [2, 1, 0.5, 0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [2, -1, -0.5, -0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [0.5, 0.5, 1, 2]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Correlation matrix:"
##      [,1] [,2] [,3] [,4]
## [1,] 1.00 0.99    0    0
## [2,] 0.99 1.00    0    0
## [3,] 0.00 0.00    1    0
## [4,] 0.00 0.00    0    1
## [1] "Coefficients: [2, 1, 0.5, 0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [2, -1, -0.5, -0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [0.5, 0.5, 1, 2]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Correlation matrix:"
##       [,1]  [,2] [,3] [,4]
## [1,]  1.00 -0.99    0    0
## [2,] -0.99  1.00    0    0
## [3,]  0.00  0.00    1    0
## [4,]  0.00  0.00    0    1
## [1] "Coefficients: [2, 1, 0.5, 0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [2, -1, -0.5, -0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [0.5, 0.5, 1, 2]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] ""
## [1] "Correlation matrix:"
##      [,1] [,2] [,3] [,4]
## [1,]    1 0.00 0.00    0
## [2,]    0 1.00 0.99    0
## [3,]    0 0.99 1.00    0
## [4,]    0 0.00 0.00    1
## [1] "Coefficients: [2, 1, 0.5, 0.5]"
## [1] "Variance of error term: = 1"
```

```
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [2, -1, -0.5, -0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [0.5, 0.5, 1, 2]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Correlation matrix:"
##      [,1]  [,2]  [,3] [,4]
## [1,]    1  0.00  0.00    0
## [2,]    0  1.00 -0.99    0
## [3,]    0 -0.99  1.00    0
## [4,]    0  0.00  0.00    1
## [1] "Coefficients: [2, 1, 0.5, 0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [2, -1, -0.5, -0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [0.5, 0.5, 1, 2]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] ""
## [1] "Correlation matrix:"
##       [,1]  [,2]   [,3]   [,4]
## [1,] 1.000 0.992  0.621  0.465
## [2,] 0.992 1.000  0.604  0.446
## [3,] 0.621 0.604  1.000 -0.177
## [4,] 0.465 0.446 -0.177  1.000
## [1] "Coefficients: [2, 1, 0.5, 0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [2, -1, -0.5, -0.5]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] "Coefficients: [0.5, 0.5, 1, 2]"
## [1] "Variance of error term: = 1"
## [1] "Variance of error term: = 25"
## [1] ""
```