

homework 2

Simon

2023-03-30

```
library(tidyverse)

## — Attaching packages — tidyverse 1.3.2 —
## ✓ ggplot2 3.4.1      ✓ purrr   1.0.1
## ✓ tibble  3.1.8      ✓ dplyr   1.1.0
## ✓ tidyr   1.3.0      ✓ stringr 1.5.0
## ✓ readr   2.1.4      ✓ forcats 1.0.0
## — Conflicts — tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()

library(MASS)

##
## 載入套件：'MASS'
##
## 下列物件被遮斷自 'package:dplyr':
##
##      select

#Q5-5
#set parameters
beta_1 <- 0.3
beta_2 <- 0.7

mu_0 <- 7
mu_1 <- 5
sigma_0 <- 1
sigma_1 <- 2
sigma_01 <- 0.5
rho_01 <- sigma_01 / (sigma_0 * sigma_1)

sigma_v <- sqrt(sigma_0 ^ 2 + sigma_1 ^ 2 - 2 * sigma_01)

#generate data
N <- 10 ^ 5

sig <- matrix(data = c(sigma_0, sigma_01, sigma_01, sigma_1), nrow = 2)
mu <- c(0, 0)
simulation_data <- as.data.frame(mvrnorm(n = N, mu = mu, Sigma = sig))
colnames(simulation_data) <- c("e0", "e1")

simulation_data <- simulation_data %>%
  mutate(X1 = sample(seq(0, 20, 1), N, replace = TRUE),
         X2 = sample(seq(0, 10, 2), N, replace = TRUE),
         w0 = mu_0 + beta_1 * X1 + e0,
         w1 = mu_1 + beta_1 * X1 +beta_2 * X2 + e1,
         D  = ifelse(w1 > w0, 1, 0),
         prop_true = 1 - pnorm((mu_0-mu_1 - beta_2*X2) / sigma_v, 0, 1))

#Q5-6
logit <- glm(D ~ X2, data = simulation_data, family = binomial)
simulation_data$prop_esti <- predict(logit, type = "response")

#Q5-7
cor(simulation_data$prop_true, simulation_data$prop_esti)

## [1] 0.9925117

#Q5-8
simulation_data <- simulation_data %>%
  mutate(weight_true = ifelse(D == 1, 1 / prop_true, 1 / (1 - prop_true)),
         weight_esti = ifelse(D == 1, 1 / prop_esti, 1 / (1 - prop_esti)),
         w = ifelse(D ==1, w1, w0))

ipw_true <- lm(w ~ D, weights = weight_true, data = simulation_data)
ipw_true$coefficients
```

```
## (Intercept)          D
##  10.297931    2.051553
```

```
ipw_esti <- lm(w ~ D, weights = weight_esti, data = simulation_data)
ipw_esti
```

```
##
## Call:
## lm(formula = w ~ D, data = simulation_data, weights = weight_esti)
##
## Coefficients:
## (Intercept)          D
##    10.404         1.715
```

```
#Q5-9
lm_w_x <- lm(w ~ X1 + X2, data = simulation_data)
summary(lm_w_x)
```

```
##
## Call:
## lm(formula = w ~ X1 + X2, data = simulation_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.3307 -0.8842 -0.0090  0.8647  6.1684
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.4951179   0.0099428   653.2  <2e-16 ***
## X1           0.3003289   0.0006781   442.9  <2e-16 ***
## X2           0.5087681   0.0012019   423.3  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.297 on 99997 degrees of freedom
## Multiple R-squared:  0.7898, Adjusted R-squared:  0.7898
## F-statistic: 1.878e+05 on 2 and 99997 DF,  p-value: < 2.2e-16
```

```
#Q5-10
lm_w0 <- lm(w ~ X1 , data = subset(simulation_data, D == 0))
summary(lm_w0)
```

```
##
## Call:
## lm(formula = w ~ X1, data = subset(simulation_data, D == 0))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.6661 -0.6643 -0.0025  0.6624  3.9786
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.1866008   0.0105494   681.2  <2e-16 ***
## X1           0.2998582   0.0009011   332.8  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9799 on 32187 degrees of freedom
## Multiple R-squared:  0.7748, Adjusted R-squared:  0.7748
## F-statistic: 1.107e+05 on 1 and 32187 DF,  p-value: < 2.2e-16
```

```
lm_w1 <- lm(w ~ X1 + X2, data = subset(simulation_data, D == 1))
summary(lm_w1)
```

```
##
## Call:
## lm(formula = w ~ X1 + X2, data = subset(simulation_data, D ==
##      1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.5477 -0.9063 -0.0190  0.8853  6.0618
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.1584728   0.0160485   383.7  <2e-16 ***
## X1           0.3003830   0.0008455   355.3  <2e-16 ***
## X2           0.5641206   0.0019014   296.7  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.33 on 67808 degrees of freedom
## Multiple R-squared:  0.7599, Adjusted R-squared:  0.7599
## F-statistic: 1.073e+05 on 2 and 67808 DF,  p-value: < 2.2e-16
```