FINC 584 PS3

15 October, 2022

Problem 1.a

```
ak_full <- readxl::read_xlsx(file.path(proj, "AK1991.xlsx"))

x_full <- matrix(data = c(rep(1,nrow(ak_full)), ak_full$edu), ncol = 2)
y_full <- matrix(data = ak_full$logwage)

beta_conventional_full <- solve(t(x_full) %*% x_full)%*%(t(x_full)%*%y_full)

resid_full <- y_full - x_full %*% beta_conventional_full
sigma2_hat_full <- mean(resid_full^2)
beta_conventional_se_full <- sqrt(diag(sigma2_hat_full *
                                         (solve(t(x_full) %*% x_full))))

kable(data.frame(beta_hat = beta_conventional_full,
                 beta_se = beta_conventional_se_full,
                 row.names = c("$\\beta_0$", "$\\beta_1$")),
      format = 'pandoc')
```

	beta_hat	beta_se
β_0	4.995182	0.0044644
β_1	0.070851	0.0003386

Problem 1.b

```
ak<- ak_full %>% slice(1:5000)

## Compute beta_hats for conventional and robust
x <- matrix(data = c(rep(1,nrow(ak)), ak$edu), ncol = 2)
y <- matrix(data = ak$logwage)
beta_conventional <- solve(t(x) %*% x)%*%(t(x)%*%y)
resid <- y - x %*% beta_conventional

## Compute SEs for conventional method
sigma2_hat <- mean(resid^2)
beta_conventional_se <- sqrt(diag(sigma2_hat * (solve(t(x) %*% x))))[2]

## Compute SEs for robust method
sandwich_bread <- solve(t(x) %*% x)
```

```r
sandwich_meat <- t(x) %*% diag(as.vector(resid)^2)  %*% x

beta_robust_se <- sqrt(diag(sandwich_bread %*% sandwich_meat %*% sandwich_bread))[2]

## Bootstrap
set.seed(0)
beta_hat_bootstrap <- rep(NA, 5000)

for(i in 1:length(beta_hat_bootstrap)){
  ak_sample <- ak[sample(1:5000, replace = T),]
  beta_hat_bootstrap[i] <- lm(logwage ~ edu, data = ak_sample)$coefficients['edu']
}

beta_bootstrap_se <- sd(beta_hat_bootstrap)

kable(data.frame(beta_hat_1 = c(beta_conventional_se, beta_robust_se,
                 beta_bootstrap_se),
                 row.names = c("Conventional", "Robust", "Bootstrap")),
      )
```

	beta_hat_1
Conventional	0.0029033
Robust	0.0029218
Bootstrap	0.0029199

```r
# ggplot(data = data.frame(beta_hat_bootstrap), aes(x = beta_hat_bootstrap)) +
#   geom_histogram(fill = 'steelblue', color = 'black') +
#   labs(title = 'Distribution of beta_hat_1') +
#   theme_bw() +
#   theme(plot.title = element_text(hjust = 0.5))
```

The results are pretty much what I expected. All three standard errors are roughly similar. The robust standard errors are larger than the conventional OLS estimators which is expected. The bootstrap method has very similar standard errors to the other two methods but is actually slightly more efficient in my sample.

Problem 1C