ECON312 Problem Set 1: question 3

Contents


```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------- tidyverse 1.2.1 --
## v ggplot2 3.2.1          v purrr   0.3.2
## v tibble  2.1.3          v dplyr   0.8.3
## v tidyr   0.8.99.9000    v stringr 1.4.0
## v readr   1.3.1          v forcats 0.4.0
## -- Conflicts ------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Monte Carlo Simulations

Consider the model:

$$Y_i = X_i'\beta + U_i$$
$$U_i|X_i \overset{i.i.d}{\sim} N(0, \sigma^2)$$

a) Define $\beta = (2, 3)^T$, $\sigma^2 = 4$; generate $N = 10,000$ values for $X \in \mathbb{R}^2$. Using your value for σ^2 draw U's

```r
set.seed(123456)
# doesn't appear that the distribution of X was specified
# I just used standard normal for x_1
N <- 10000


X_0 <- rep(1, N)
X_1 <-rnorm(n = N)

sigma <- 2
U <- rnorm(n = N, sd = sigma)

data <- tibble(X_0 = X_0,
               X_1 = X_1,
               U = U)

knitr::kable(head(data))
```

X_0	X_1	U
1	0.8337332	1.4791073
1	-0.2760478	3.5651206
1	-0.3550018	-3.0699699
1	0.0874874	0.0054147
1	2.2522557	0.6170447
1	0.8344601	4.3414702

Finally, compute the Y's

```
beta <- c(2,3)

data <- data %>%
  mutate(Y = X_0*beta[[1]] + X_1*beta[[2]] + U)

knitr::kable(head(data))
```

X_0	X_1	U	Y
1	0.8337332	1.4791073	5.980307
1	-0.2760478	3.5651206	4.736977
1	-0.3550018	-3.0699699	-2.134975
1	0.0874874	0.0054147	2.267877
1	2.2522557	0.6170447	9.373812
1	0.8344601	4.3414702	8.844851

Estimate $\hat{\beta}$ and its standard errors from your data using standard OLS formulas.

We did the actual matrix calculations

$$\hat{beta} = (XX')^{-1}(XY)$$

```
X <- as.matrix(tibble(int =1,
            X_1 = data$X_1))

Y <- data$Y

beta_n <- solve(t(X)%*%X)%*%t(X)%*%Y

beta_n
```

```
##          [,1]
## int 2.012333
## X_1 2.962278
```

Standard errors

Assuming homoskedasticity,

$$V = XX'\hat{\sigma}^2$$

$$se(\hat{\beta}_k) = \sqrt{\frac{1}{n} diag(\hat{V})_k}$$

```r
u <- Y - X%*%beta_n

u_sq <- as.vector(u *u)

sigma_sq_hat <- sum(u_sq)/N

V <- solve(t(X)%*%X)*sigma_sq_hat

se <- sqrt(diag(V))

se
```

```
##        int        X_1
## 0.02000305 0.02003243
```

Verifying with statistical software

```r
ols <- lm(Y ~ X_1, data)

summary(ols)
```

```
##
## Call:
## lm(formula = Y ~ X_1, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5376 -1.3689 -0.0049  1.3556  7.5141
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.01233    0.02001   100.6   <2e-16 ***
## X_1          2.96228    0.02003   147.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2 on 9998 degrees of freedom
## Multiple R-squared:  0.6862, Adjusted R-squared:  0.6862
## F-statistic: 2.186e+04 on 1 and 9998 DF,  p-value: < 2.2e-16
```