

Simulations and Bias of LS estimator

Loading data from previous task

```
library(faraway)
data("uswages")
attach(uswages)
```

Obtaining the necessary variables and parameters

```
n = nrow(uswages)
sigma = 400
mu = -250 + 50 * educ + 10 * exper
c = (mu - min(mu)) / (max(mu) - min(mu))
```

The settings for residuals

```
# eps have to be functions to produce different results every iteration
eps.1 = function() rnorm(n, 0, sigma)
eps.2 = function() rnorm(n, 0, c * sigma)
eps.3 = function() rnorm(n, 1000 * sin(mu), sigma)
```

We expect that non-central residuals will introduce bias in the LS estimator.

Simulation

```
Nsim = 1000
simulate = function(eps) {
  estimators = matrix(NA, nrow = Nsim, ncol = 3)
  for(sim in 1:Nsim) {
    Y = mu + eps()
    fit = lm(Y ~ educ + exper)
    estimators[sim,] = coef(fit)
  }
  return(colMeans(estimators))
}

simulate(eps.1)
```

```
[1] -249.909827  50.064566   9.978591
```

```
simulate(eps.2)
```

```
[1] -250.00763  50.01100  10.00507
```

```
simulate(eps.3)
```

```
[1] -270.74833  46.21445  14.32269
```

We can see eps.3 introduces some bias though it's not that big when we take into consideration the oscillating nature of `sin(.)` - The LS estimators are biased, but in the mean they cancel out.