Problem Set 1

Owen Jetton

04/06/2022

Question 1

```
data = read.csv("data-001.csv")
```

Question 2

```
reg1 = lm(data = data,
          formula = income_black_2010 ~ pop_enslaved_1860 + pop_total_1860 + pop_total_2010)

q2_coef = reg1$coefficients[["pop_enslaved_1860"]]
```

The coefficient on "pop_enslaved_1980" is -0.2670247

Question 3

```
# endogenous variable
Y = as.matrix(data$income_black_2010)

# exogenous variables (with intercept)
X = matrix(c(rep(1, 710), data$pop_enslaved_1860, data$pop_total_1860, data$pop_total_2010),
            ncol = 4)

reg_q3 = solve(t(X) %*% X) %*% t(X) %*% Y

reg_q3[2,1]
```

```
## [1] -0.2670247
```

The coefficient on "pop_enslaved_1980" is `rreg_q3[2,1]`' which is the same as in question 2.

Question 4

```
reg_fun = function(y, x) {
  coef = solve(t(x) %*% x) %*% t(x) %*% y

  return(coef)
}
```

Question 5

```r
reg_fun2 = function(y, x) {

  # coefficient equation
  coef = solve(t(x) %*% x) %*% t(x) %*% y

  # standard errors
  e = (y - x %*% coef)
  s_sq = (1/(dim(x)[1] - dim(x)[2]-1))*sum(e^2)

  # calculate variance matrix
  variance_matrix = s_sq * solve(t(x) %*% x)

  # arrange the results
  stnd_errors = sqrt(diag(variance_matrix))

  return(stnd_errors)

}

reg_fun2(Y, X)
```

```
## [1] 6.892133e+02 1.292888e-01 6.353467e-02 1.804909e-03
```