Manual Regression

Part I — Generating Data

1. A. Generate variables X1 and X2, each containing 20 numbers where X1 ~ U (25, 30), X2 ~ U (40, 50)!

Program:

```
x1 <- runif(20, min = 25, max = 30)
x2 <- runif(20, min = 40, max = 50)
```

output:

```
> x1
 [1] 29.62592 27.66287 26.44799 25.92575 27.39397 27.29407 28.64658 28.22063 27.67092 27.47719 26.46824
[12] 25.51077 26.52225 28.01369 28.82644 27.37414 25.91507 26.14232 26.70942 28.07801
> x2
 [1] 47.68492 45.43451 44.08556 48.02206 46.75533 42.88051 40.71887 43.22962 45.79145 44.38357 45.59294
[12] 48.21842 46.12673 43.07281 41.08253 47.67222 47.26494 45.72463 46.33105 40.65968
```

1. B. Generate e, data error, where E ~ N(0,1)!

Program:

```
e <- rnorm(20, mean = 0, sd = 1)
```

output:

```
> e
 [1] -1.31813162 -0.64371002 -0.42377824  0.15334479 -0.51890895 -0.16877663  1.02416155 -0.59318087
 [9] -1.01059764 -1.38272581  2.51690205 -1.57524348 -0.00240107 -1.01567876  1.10662101  0.09259968
[17] -0.07604643  0.19100380 -0.20615660 -0.07824785
```

1. C. Generate Y following the equation $Y = 2 + 1.5X1 + 4X2 + E$!

Program:

```
y <- 2+1.5*x1+a*x2+e
```

output:

```
> y
 [1] 235.8604 224.5886 217.5905 233.1302 229.5934 214.2944 208.8695 216.6562 225.6616 219.3673 226.5910
[12] 231.5646 226.2879 215.2961 210.6764 233.8427 229.8563 224.3030 227.1822 206.6775
```

1. D. Create a data frame using generated data. This data frame will then be used to test the function.

Program:

```
datatest <- cbind(y,x1,x2)
```

Part II — Creating Function

Program:

```
manualregression <- function(regdata, alpha = 0.05){
 #dataprep
 Y <- as.matrix(regdata[,1])
 n <- length(Y)
 m <- dim(regdata[,-1])[2]
 onesvect <- rep(1,n)
 X <- cbind(onesvect,regdata[,-1])
 onesmatr <- matrix(rep(1,n*n),n,n)
 #works
```

```
 #coefficients
 B <- solve(t(X)%*%X)%*%t(X)%*%Y
 rownames(B) <- c('b0','b1','b2')
 colnames(B) <- 'Coefficients'
 #anova table
 #sources
 sources <- c('regression','error','total')
 #df
 df <- c(m,n-m-1,n-1)
 #ss
 ssr <- as.numeric(t(B)%*%t(X)%*%Y-(1/n)*t(Y)%*%onesmatr%*%Y)
 sse <- as.numeric(t(Y)%*%Y-t(B)%*%t(X)%*%Y)
 ssto <- ssr + sse
 ss <- c(ssr,sse,ssto)
 #ms
 msr <- ssr/df[1]
 mse <- sse/df[2]
 ms <- c(msr,mse,NA)
 #f
 f <- c(msr/mse, NA, NA)
 pvalue <- c(1-pf(msr/mse, m, n-m-1), NA, NA)
 anovatable <- data.frame(sources,df,ss,ms,f,pvalue)
 #hypothesis testing
 h0 <- 'There is no relationship between Y and X variables'
 h1 <- 'There is a relationship between Y and at least one of the X variables'
 tablef <- qf(1-alpha,m,n-m-1)
 if(tablef<msr/mse){
   hypothesis <- h1
 } else{
   hypothesis <- h0
 }
 #regression line
 if(hypothesis==h1){
   regline <- sprintf('The regression line, therefore, is Y = %s + %sX1 + %sX2.', B[1], B[2], B[3])
 } else{
   regline <- 'Since there is no relationship between Y and any of the X variables, the
regression line produced is not significant.'
 }
 results <- list(B, anovatable, hypothesis, regline)
 return(results)
}
#test
manualregression(datatest)
```

output:

```
> manualregression(datatest)
[[1]]
   Coefficients
b0    16.709131
b1     1.238086
b2     3.827778

[[2]]
     sources df         ss          ms        f pvalue
1 regression  2 1422.58830 711.2941486 811.1661      0
2      error 17   14.90694   0.8768786       NA     NA
3      total 19 1437.49523          NA       NA     NA

[[3]]
[1] "There is a relationship between Y and at least one of the X variables"

[[4]]
[1] "The regression line, therefore, is Y = 16.709131419787 + 1.23808611275546X1 + 3.82777801247837X2."
```

Regresi Manual

Bagian I — Persiapan Data

1. A. Bangkitkan bilangan X1 dan X2 sebanyak 20 bilangan, X1 ~ U (25, 30), X2 ~ U (40, 50)

Program:

```
x1 <- runif(20, min = 25, max = 30)
x2 <- runif(20, min = 40, max = 50)
```

Hasil:

```
> x1
 [1] 29.62592 27.66287 26.44799 25.92575 27.39397 27.29407 28.64658 28.22063 27.67092 27.47719 26.46824
[12] 25.51077 26.52225 28.01369 28.82644 27.37414 25.91507 26.14232 26.70942 28.07801
> x2
 [1] 47.68492 45.43451 44.08556 48.02206 46.75533 42.88051 40.71887 43.22962 45.79145 44.38357 45.59294
[12] 48.21842 46.12673 43.07281 41.08253 47.67222 47.26494 45.72463 46.33105 40.65968
```

1. B. Bangkitkan bilangan E ~ N (0, 1) sebanyak 20 bilangan.

Program:

```
e <- rnorm(20, mean = 0, sd = 1)
```

Hasil:

```
> e
 [1] -1.31813162 -0.64371002 -0.42377824  0.15334479 -0.51890895 -0.16877663  1.02416155 -0.59318087
 [9] -1.01059764 -1.38272581  2.51690205 -1.57524348 -0.00240107 -1.01567876  1.10662101  0.09259968
[17] -0.07604643  0.19100380 -0.20615660 -0.07824785
```

2. C. Bangkitkan bilangan Y dengan formula $Y = 2 + 1.5X1 + 4X2 + E$!

Program:

```
y <- 2+1.5*x1+a*x2+e
```

Hasil:

```
> y
 [1] 235.8604 224.5886 217.5905 233.1302 229.5934 214.2944 208.8695 216.6562 225.6616 219.3673 226.5910
[12] 231.5646 226.2879 215.2961 210.6764 233.8427 229.8563 224.3030 227.1822 206.6775
```

1. D. Bentuk data menjadi dataframe untuk menguji fungsi.

Program:

```
datatest <- cbind(y,x1,x2)
```

Bagian II — Fungsi

Program:

```
manualregression <- function(regdata, alpha = 0.05){
 #dataprep
 Y <- as.matrix(regdata[,1])
 n <- length(Y)
 m <- dim(regdata[,-1])[2]
 onesvect <- rep(1,n)
 X <- cbind(onesvect,regdata[,-1])
 onesmatr <- matrix(rep(1,n*n),n,n)
 #works
 #coefficients
 B <- solve(t(X)%*%X)%*%t(X)%*%Y
```

```
 rownames(B) <- c('b0','b1','b2')
 colnames(B) <- 'Coefficients'
 #anova table
 #sources
 sources <- c('regression','error','total')
 #df
 df <- c(m,n-m-1,n-1)
 #ss
 ssr <- as.numeric(t(B)%*%t(X)%*%Y-(1/n)*t(Y)%*%onesmatr%*%Y)
 sse <- as.numeric(t(Y)%*%Y-t(B)%*%t(X)%*%Y)
 ssto <- ssr + sse
 ss <- c(ssr,sse,ssto)
 #ms
 msr <- ssr/df[1]
 mse <- sse/df[2]
 ms <- c(msr,mse,NA)
 #f
 f <- c(msr/mse, NA, NA)
 pvalue <- c(1-pf(msr/mse, m, n-m-1), NA, NA)
 anovatable <- data.frame(sources,df,ss,ms,f,pvalue)
 #hypothesis testing
 h0 <- 'There is no relationship between Y and X variables'
 h1 <- 'There is a relationship between Y and at least one of the X variables'
 tablef <- qf(1-alpha,m,n-m-1)
 if(tablef<msr/mse){
   hypothesis <- h1
 } else{
   hypothesis <- h0
 }
 #regression line
 if(hypothesis==h1){
   regline <- sprintf('The regression line, therefore, is Y = %s + %sX1 + %sX2.', B[1], B[2], B[3])
 } else{
   regline <- 'Since there is no relationship between Y and any of the X variables, the
regression line produced is not significant.'
 }
 results <- list(B, anovatable, hypothesis, regline)
 return(results)
}
#test
manualregression(datatest)
```

Hasil:

```
> manualregression(datatest)
[[1]]
   Coefficients
b0    16.709131
b1     1.238086
b2     3.827778

[[2]]
     sources df         ss          ms        f pvalue
1 regression  2 1422.58830 711.2941486 811.1661      0
2      error 17   14.90694   0.8768786       NA     NA
3      total 19 1437.49523          NA       NA     NA

[[3]]
[1] "There is a relationship between Y and at least one of the X variables"

[[4]]
[1] "The regression line, therefore, is Y = 16.709131419787 + 1.23808611275546X1 + 3.82777801247837X2."
```