

Multiple Regression Model

Thach Pham

15/11/2018

```
library(wooldridge)
library(tidyverse)
library(stargazer)
library(GGally)
```

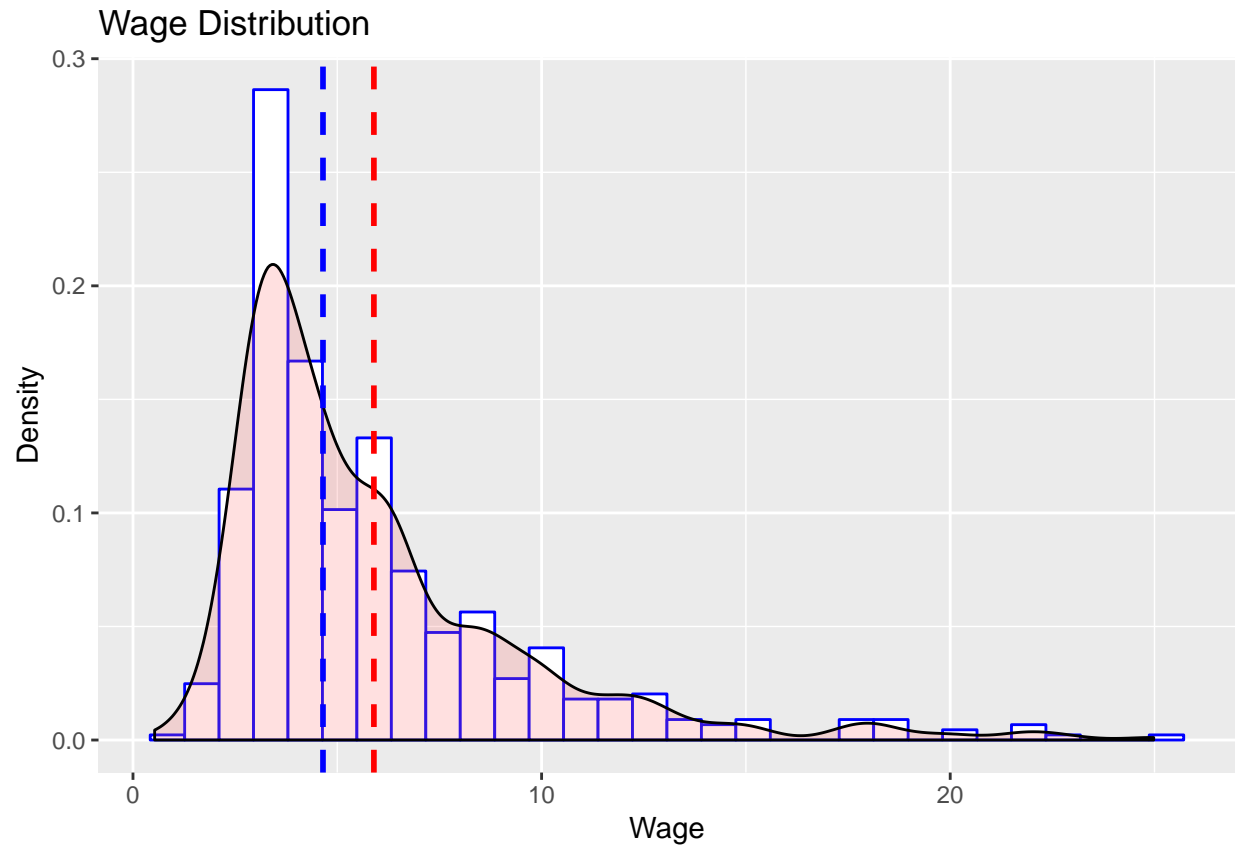
Descriptive Statistics of wage1 data

```
data(wage1)
desc.wage1 <- select(wage1, c(wage, educ, exper, tenure, lwage))
stargazer(desc.wage1, type = "text", style = "qje", title = "Descriptive Statistics",
           digits = 4, out = "MRO/Descriptive Statistics.doc")
```

```
##
## Descriptive Statistics
## -----
## Statistic   N    Mean    St. Dev.    Min    Pctl(25) Pctl(75)    Max
## =====
## wage        526  5.8961    3.6931    0.5300    3.3300    6.8800    24.9800
## educ        526 12.5627    2.7690     0         12         14         18
## exper       526 17.0171   13.5722     1          5         26         51
## tenure      526  5.1046    7.2245     0          0          7         44
## lwage       526  1.6233    0.5315   -0.6349    1.2030    1.9286    3.2181
## =====
```

Distribution of Wage

```
(wage.dist <- ggplot(desc.wage1, aes(wage)) +
  geom_histogram(aes(y = ..density..), binwidth = 0.5, colour = "blue", fill = "white") +
  geom_density(alpha = 0.2, fill = "#FF6666") +
  labs(x = "Wage", y = "Density", title = "Wage Distribution") +
  geom_vline(aes(xintercept = mean(wage)), color = "red", linetype = "dashed", size = 1) +
  geom_vline(aes(xintercept = median(wage)), color = "blue", linetype = "dashed", size = 1))
```



```
ggsave("MRO/Wage Distribution.png", wage.dist)
```

Distribution of Log Wage

```
(lwage.dist <- ggplot(desc.wage1, aes(lwage)) +
  geom_histogram(aes(y = ..density..), bandwidth = 0.5, colour = "blue", fill = "white") +
  geom_density(alpha = 0.2, fill = "#FF6666") +
  labs(x = "Log Wage", y = "Density", title = "Log Wage Distribution") +
  geom_vline(aes(xintercept = mean(lwage)), color = "red", linetype = "dashed", size = 1) +
  geom_vline(aes(xintercept = median(lwage)), color = "blue", linetype = "dashed", size = 1))
```

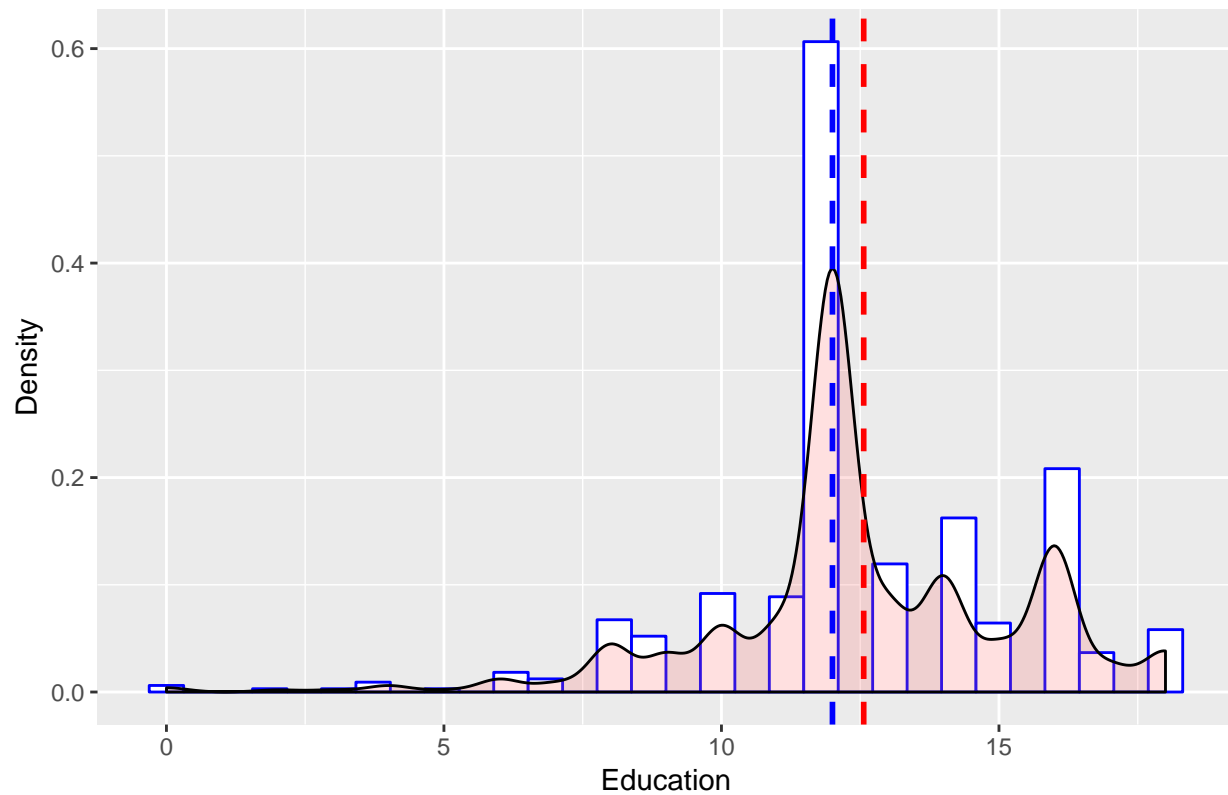


```
ggsave("MRO/Log Wage Distribution.png", lwage.dist)
```

Distribution of Education

```
(educ.dist <- ggplot(desc.wage1, aes(educ)) +  
  geom_histogram(aes(y = ..density..), bandwidth = 0.5, colour = "blue", fill = "white") +  
  geom_density(alpha = 0.2, fill = "#FF6666") +  
  labs(x = "Education", y = "Density", title = "Education Distribution") +  
  geom_vline(aes(xintercept = mean(educ)), color = "red", linetype = "dashed", size = 1) +  
  geom_vline(aes(xintercept = median(educ)), color = "blue", linetype = "dashed", size = 1))
```

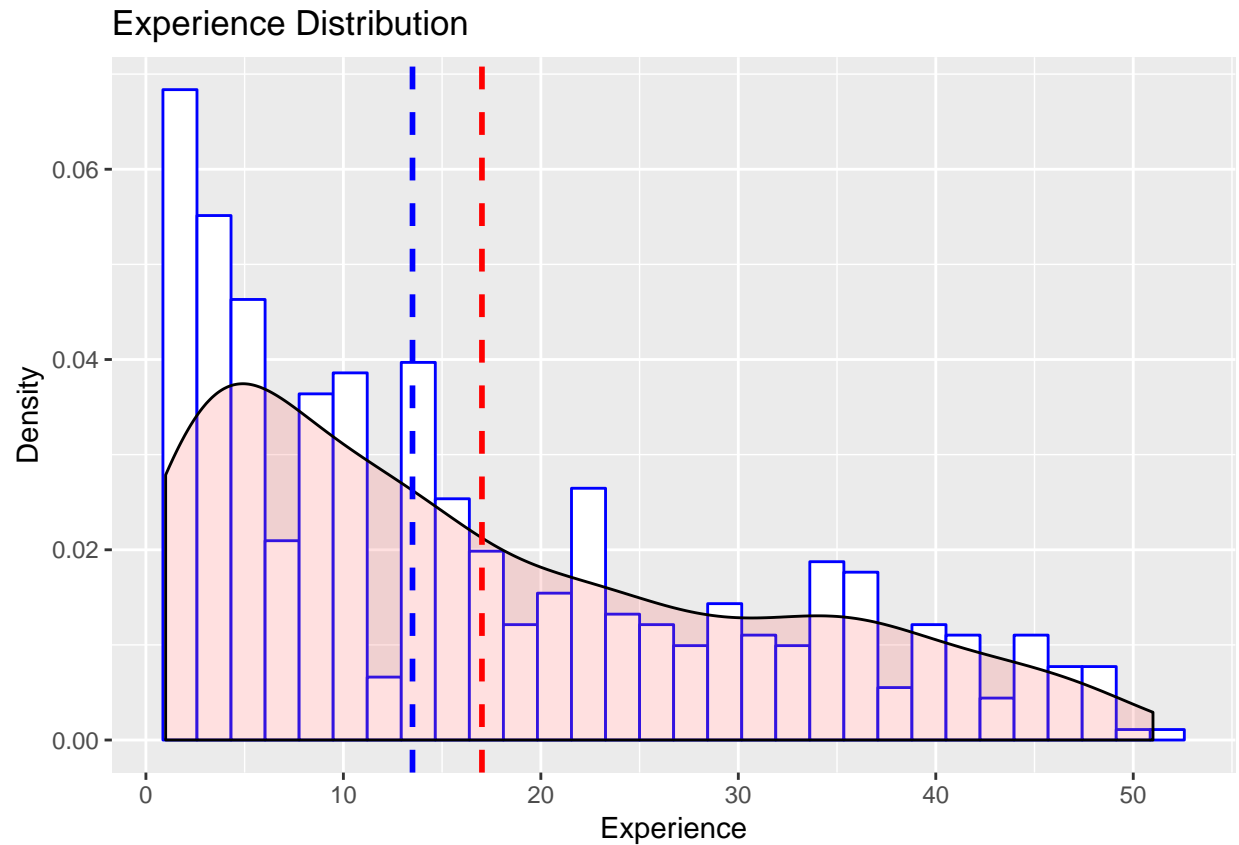
Education Distribution



```
ggsave("MRO/Education Distribution.png", educ.dist)
```

Distribution of Experience

```
(exper.dist <- ggplot(desc.wage1, aes(exper)) +  
  geom_histogram(aes(y = ..density..), bandwidth = 0.5, colour = "blue", fill = "white") +  
  geom_density(alpha = 0.2, fill = "#FF6666") +  
  labs(x = "Experience", y = "Density", title = "Experience Distribution") +  
  geom_vline(aes(xintercept = mean(exper)), color = "red", linetype = "dashed", size = 1) +  
  geom_vline(aes(xintercept = median(exper)), color = "blue", linetype = "dashed", size = 1))
```



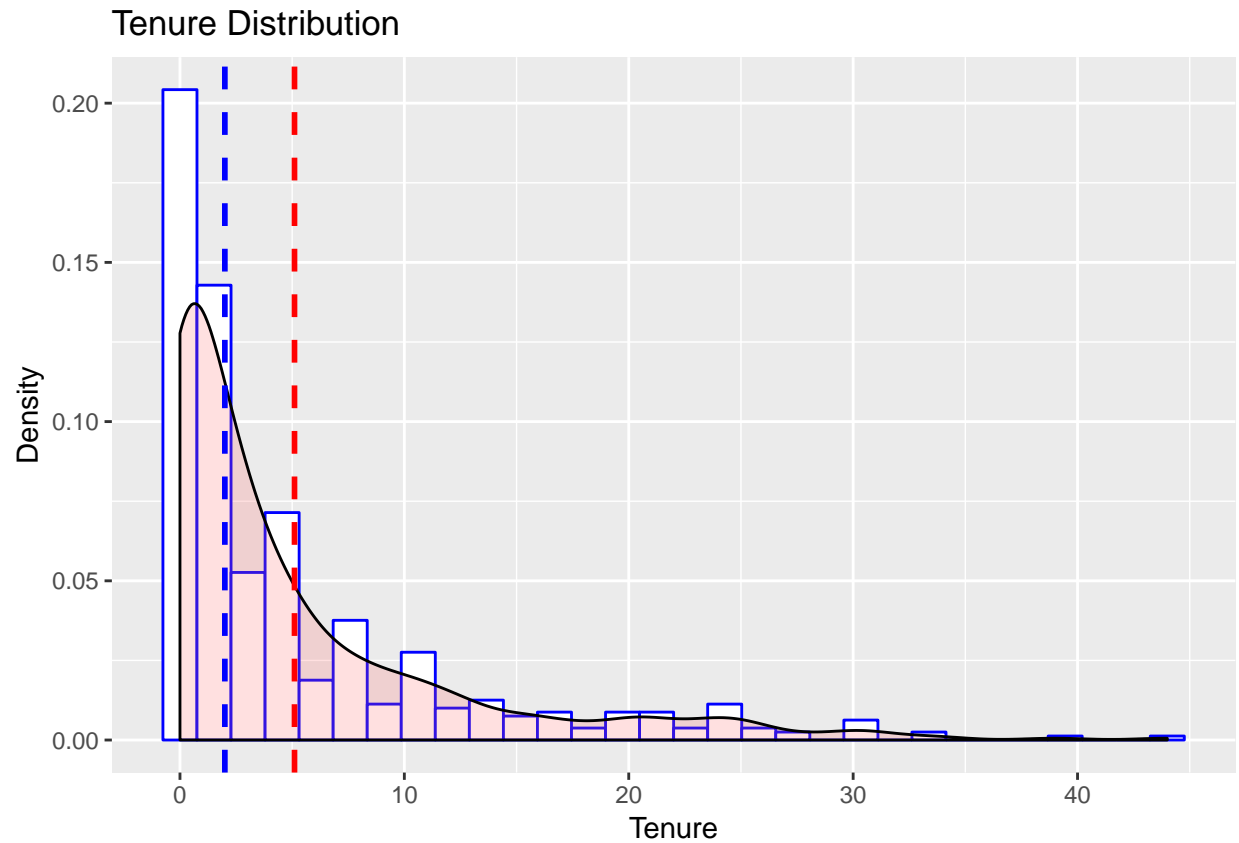
```
ggsave("MRO/Experience Distribution.png", exper.dist)
```

Distribution of Tenure

```
(tenure.dist <- ggplot(desc.wage1, aes(tenure)) +
  geom_histogram(aes(y = ..density..), binwidth = 0.5, colour = "blue", fill = "white") +
  geom_density(alpha = 0.2, fill = "#FF6666") +
  labs(x = "Tenure", y = "Density", title = "Tenure Distribution") +
  geom_vline(aes(xintercept = mean(tenure)), color = "red", linetype = "dashed", size = 1) +
  geom_vline(aes(xintercept = median(tenure)), color = "blue", linetype = "dashed", size = 1))
```

```
## Warning: Ignoring unknown parameters: binwidth
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
ggsave("MRO/Tenure Distribution.png", tenure.dist)
```

```
## Saving 6.5 x 4.5 in image
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Covariance and Correlation Matrix

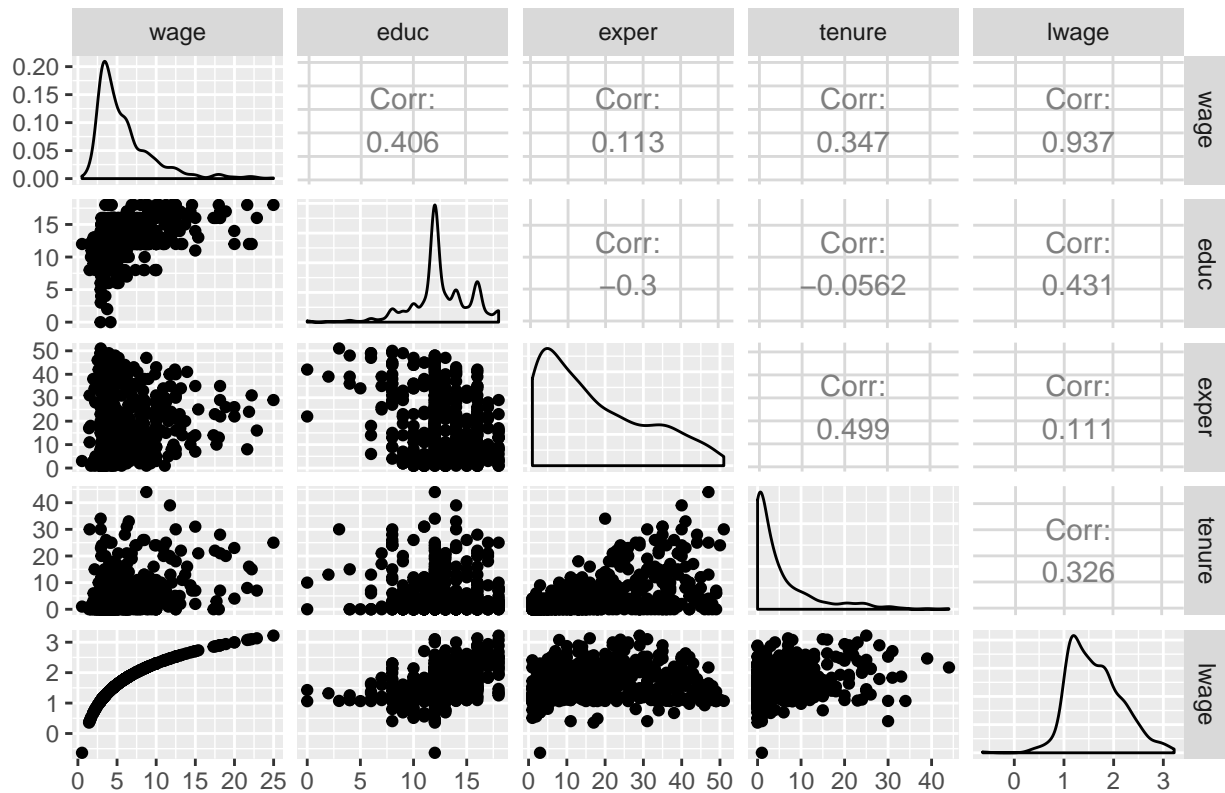
```
cov(desc.wage1)
```

```
##           wage           educ           exper           tenure           lwage
## wage  13.638884   4.1508640   5.6590763   9.255208  1.8394674
## educ   4.150864   7.6674851  -11.2572660  -1.123715  0.6344412
## exper   5.659076 -11.2572660  184.2035162  48.956303  0.8034574
## tenure  9.255208  -1.1237154  48.9563027  52.192855  1.2500910
## lwage   1.839467   0.6344412   0.8034574   1.250091  0.2825329
```

Correlation Matrix

```
ggpairs(desc.wage1, title = "Correlation Matrix")
```

Correlation Matrix



Regression Models

```
# Model 1
modell1 <- lm(wage ~ educ + exper + tenure, data = desc.wage1)
stargazer(modell1, type = "text", title = "lm(wage ~ educ + exper + tenure",
           style = "qje", out = "MRO/modell1.doc")
```

```
##
## lm(wage ~ educ + exper + tenure
## =====
##                                wage
## -----
## educ                          0.599***
##                               (0.051)
##
## exper                          0.022*
##                               (0.012)
##
## tenure                        0.169***
```

```

##                                (0.022)
##
## Constant                      -2.873***
##                                (0.729)
##
## N                             526
## R2                           0.306
## Adjusted R2                   0.302
## Residual Std. Error          3.084 (df = 522)
## F Statistic                   76.873*** (df = 3; 522)
## =====
## Notes:                        ***Significant at the 1 percent level.
##                               **Significant at the 5 percent level.
##                               *Significant at the 10 percent level.

```

```
# Model 2
```