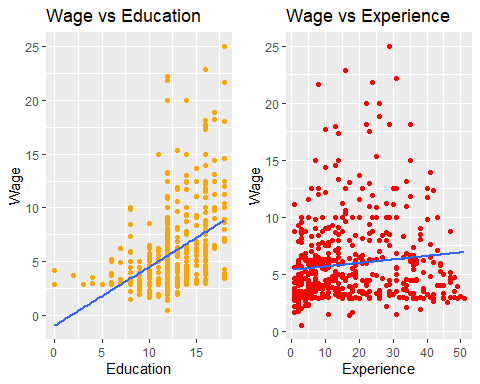
#NORMALITY TESTS IN ECONOMETRICS USING R  
##Author: R J Neel  
###Referene: Ben Lambert Econometrics Course, Introductory Econometrics: A modern approach by J Wooldrige, Econometric Analysis by W. Greene

Let us consider an example from the **wage1** dataset. Here, we test the relationship between wage and education and experience.

## wage educ exper tenure nonwhite female married numdep smsa northcen south  
## 1 3.10 11 2 0 0 1 0 2 1 0 0  
## 2 3.24 12 22 2 0 1 1 3 1 0 0  
## 3 3.00 11 2 0 0 0 0 2 0 0 0  
## 4 6.00 8 44 28 0 0 1 0 1 0 0  
## 5 5.30 12 7 2 0 0 1 1 0 0 0  
## 6 8.75 16 9 8 0 0 1 0 1 0 0  
## west construc ndurman trcommpu trade services profserv profocc clerocc  
## 1 1 0 0 0 0 0 0 0 0  
## 2 1 0 0 0 0 1 0 0 0  
## 3 1 0 0 0 1 0 0 0 0  
## 4 1 0 0 0 0 0 0 0 1  
## 5 1 0 0 0 0 0 0 0 0  
## 6 1 0 0 0 0 0 1 1 0  
## servocc lwage expersq tenursq  
## 1 0 1.131402 4 0  
## 2 1 1.175573 484 4  
## 3 0 1.098612 4 0  
## 4 0 1.791759 1936 784  
## 5 0 1.667707 49 4  
## 6 0 2.169054 81 64



We devise the following model:  
wage=f(education, experience)

Therefore, we devise the following model:

The result of regression is

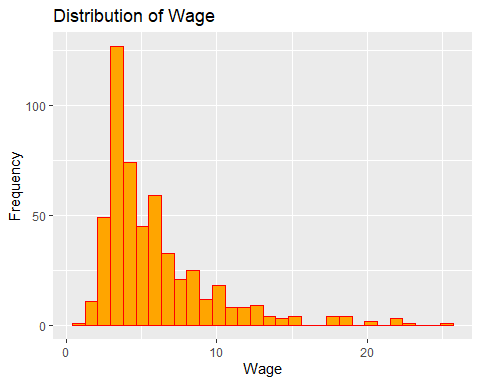
##   
## Call:  
## lm(formula = wage ~ educ + exper, data = data)  
##   
## Coefficients:  
## (Intercept) educ exper   
## -3.3905 0.6443 0.0701

##   
## Call:  
## lm(formula = wage ~ educ + exper, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.5532 -1.9801 -0.7071 1.2030 15.8370   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.39054 0.76657 -4.423 1.18e-05 \*\*\*  
## educ 0.64427 0.05381 11.974 < 2e-16 \*\*\*  
## exper 0.07010 0.01098 6.385 3.78e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.257 on 523 degrees of freedom  
## Multiple R-squared: 0.2252, Adjusted R-squared: 0.2222   
## F-statistic: 75.99 on 2 and 523 DF, p-value: < 2.2e-16

## Analysis of Variance Table  
##   
## Response: wage  
## Df Sum Sq Mean Sq F value Pr(>F)   
## educ 1 1179.7 1179.73 111.208 < 2.2e-16 \*\*\*  
## exper 1 432.5 432.52 40.772 3.78e-10 \*\*\*  
## Residuals 523 5548.2 10.61   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

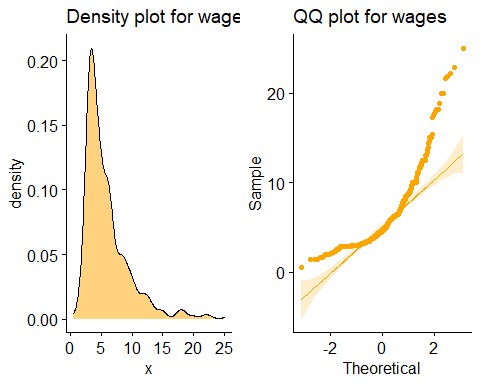
Thus, we have

Now, we test for **NORMALITY**

Let us see the distribution of wages  


Prima facie it looks like the distribution is **skewed** towards the right.

Let us conduct all the normality tests

**[1]Visual Tests**  


**[2]Jarque-Berra Test**

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

##   
## Jarque Bera Test  
##   
## data: data$wage  
## X-squared = 894.62, df = 2, p-value < 2.2e-16

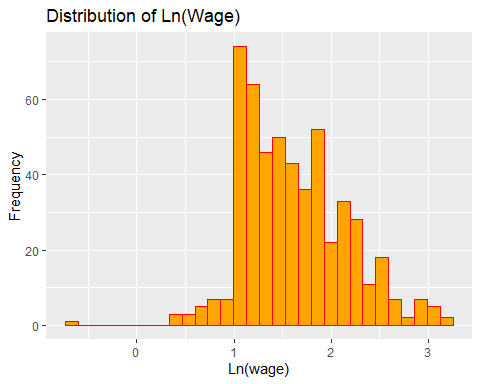
**INTERPRETATION**  
For Jarque-Berra test our  
*Ho: Distribution is NORMAL*  
*Ha: Otherwise*  
A p value of implies That the test stastic lies in the REJECTION REGION. Therefore, We **fail** to conclude that the distribution is **NORMAL**. It was already hinted in the Visual tests.

**[3]Shapiro-Wilk Test**

##   
## Shapiro-Wilk normality test  
##   
## data: data$wage  
## W = 0.80273, p-value < 2.2e-16

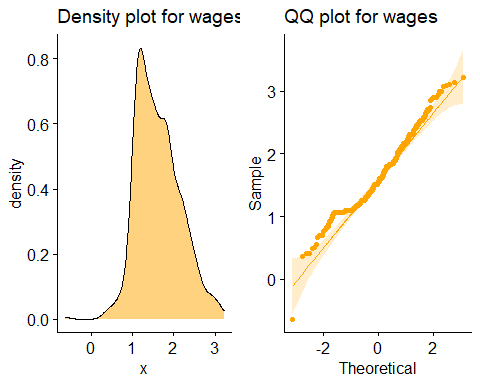
As is in the case for Jarque-Berra test, In Shapiro-Wilkins test too = our  
*Ho: Distribution is NORMAL*  
*Ha: Otherwise*  
A p value of implies That the test stastic lies in the REJECTION REGION. Therefore, We **fail** to conclude that the distribution is **NORMAL**. It was already hinted in the Visual tests.

Shapiro-wilk is the more preferred test.

**Let us now repeat the exercise with Log(wages)**  
Let us see the distribution of Log(wages)  


Prima facie it does **NOT** look **NORMAL**

Let us conduct all the normality tests

**[1]Visual Tests**  


**[2]Jarque-Berra Test**

##   
## Jarque Bera Test  
##   
## data: log(data$wage)  
## X-squared = 16.67, df = 2, p-value = 0.0002399

**INTERPRETATION**  
The p value has increased but it is still implies That the test stastic lies in the REJECTION REGION. Therefore, We **fail** to conclude that the distribution is **NORMAL**. It was already hinted in the Visual tests.

**[3]Shapiro-Wilk Test**

##   
## Shapiro-Wilk normality test  
##   
## data: log(data$wage)  
## W = 0.96909, p-value = 4.423e-09

As is in the case for Jarque-Berra test, In Shapiro-Wilkins test too it fails to prove **NORMALITY**