

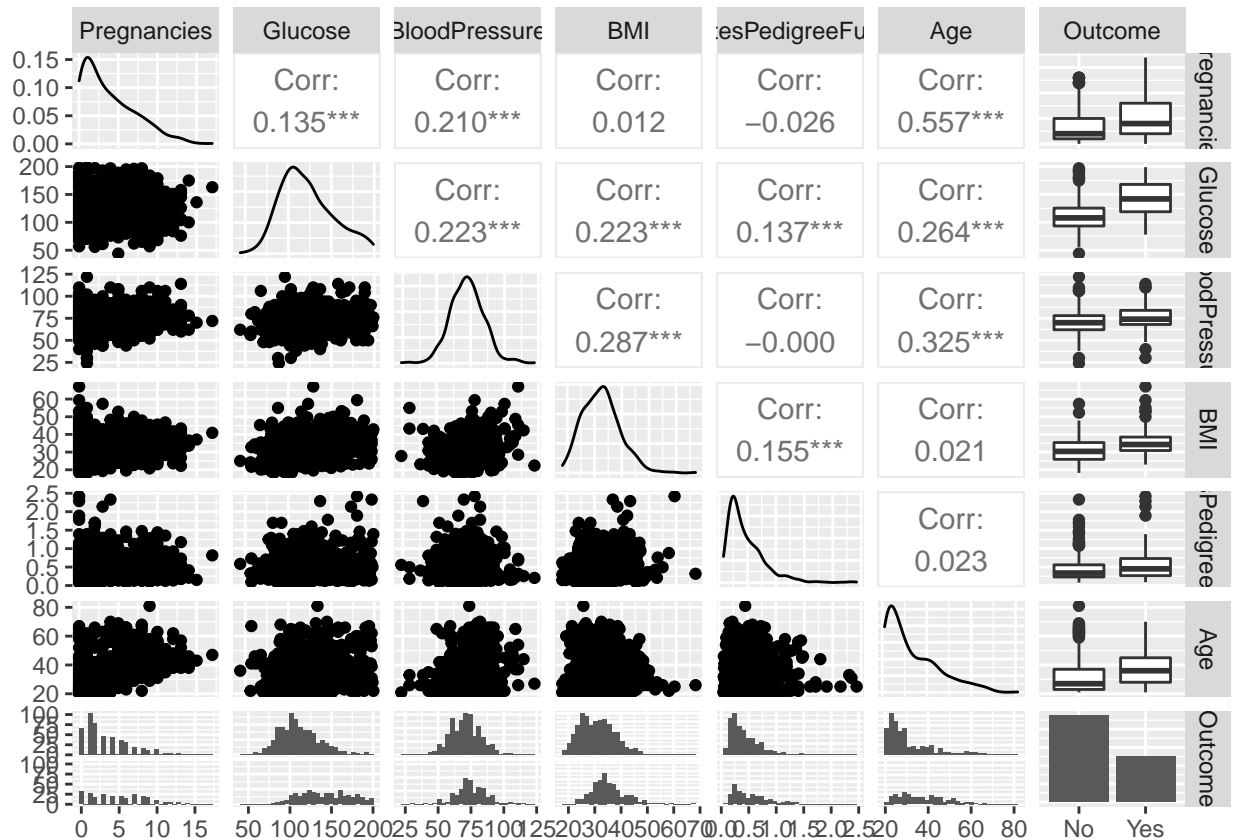
Multiple Linear Regression Runs

Uyen Nguyen

Last iteration of MLR

Scatterplots

```
GGally::ggpairs(diabetes2)
```



Full MLR with Pregnancies, Blood Pressure, BMI, DiabetesPedigreeFunction, and Age

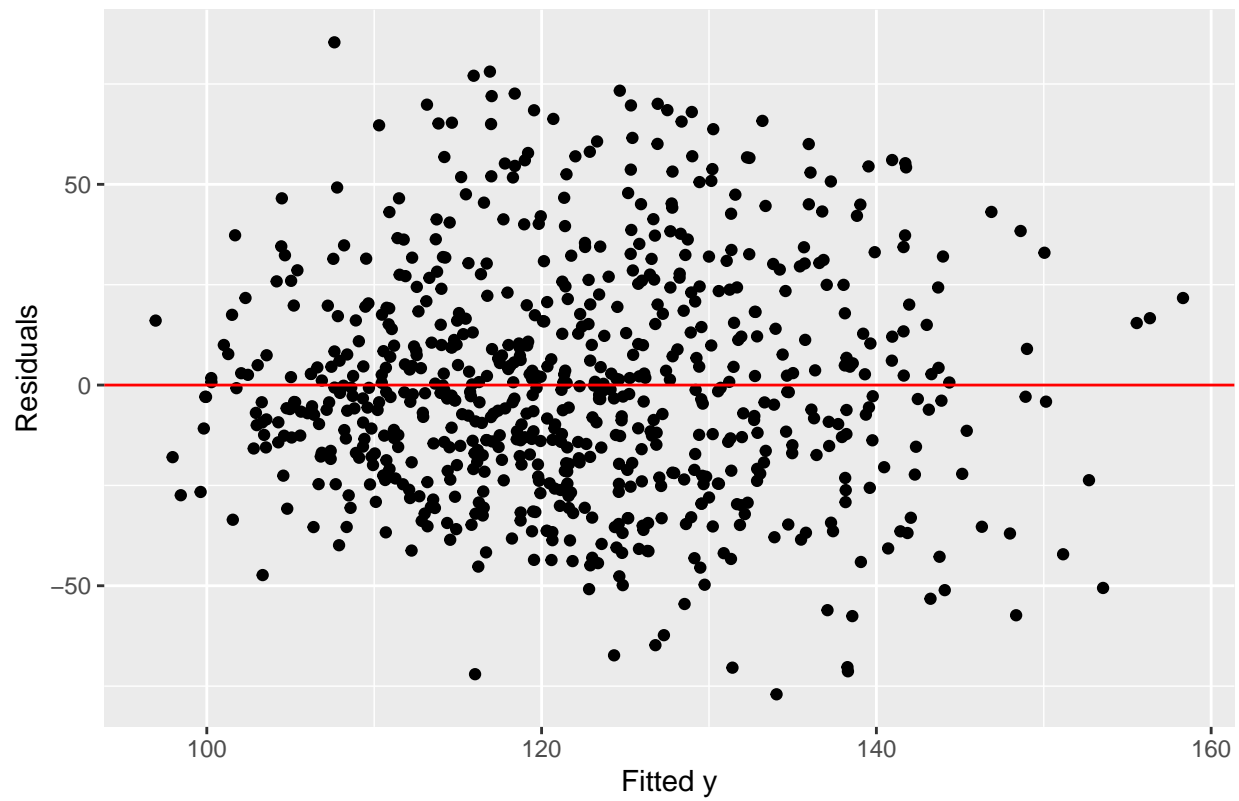
```
result <- lm(Glucose ~ Pregnancies + BloodPressure + BMI + DiabetesPedigreeFunction + Age, data = diabetes2)
summary(result)
```

```
##
## Call:
## lm(formula = Glucose ~ Pregnancies + BloodPressure + BMI + DiabetesPedigreeFunction +
##      Age, data = diabetes2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -77.037 -19.789  -2.746  17.503  85.378
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    54.22498     7.39155   7.336 5.96e-13 ***
## Pregnancies    -0.14434     0.38287  -0.377  0.70629
## BloodPressure    0.25063     0.09568   2.619  0.00899 **
## BMI             0.77448     0.16452   4.707 3.01e-06 ***
## DiabetesPedigreeFunction 9.62002     3.26048   2.950  0.00328 **
## Age            0.61042     0.11342   5.382 9.98e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.7 on 718 degrees of freedom
## Multiple R-squared:  0.1349, Adjusted R-squared:  0.1289
## F-statistic: 22.39 on 5 and 718 DF, p-value: < 2.2e-16
```

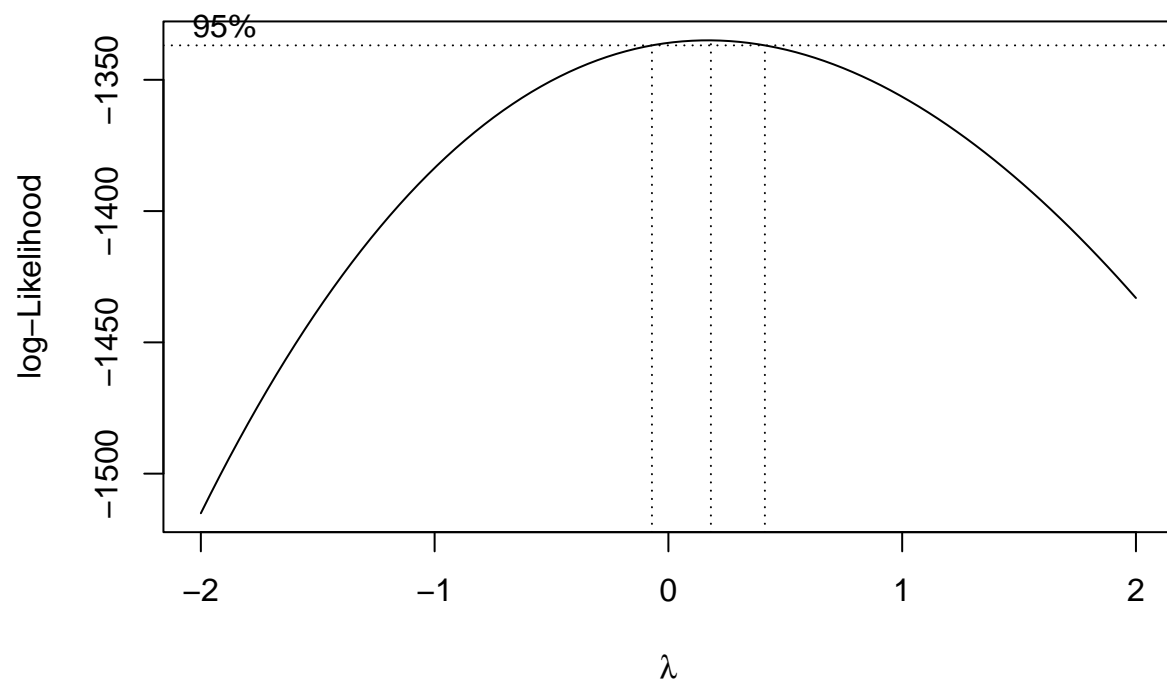
```
yhat <- result$fitted.values
res <- result$residuals
diabetes2 <- data.frame(diabetes2, yhat, res)

ggplot(diabetes2, aes(x = yhat, y = res))+
  geom_point()+
  geom_hline(yintercept=0, color="red")+
  labs(x="Fitted y",
       y="Residuals",
       title="Residual Plot from full MLR")
```

Residual Plot from full MLR

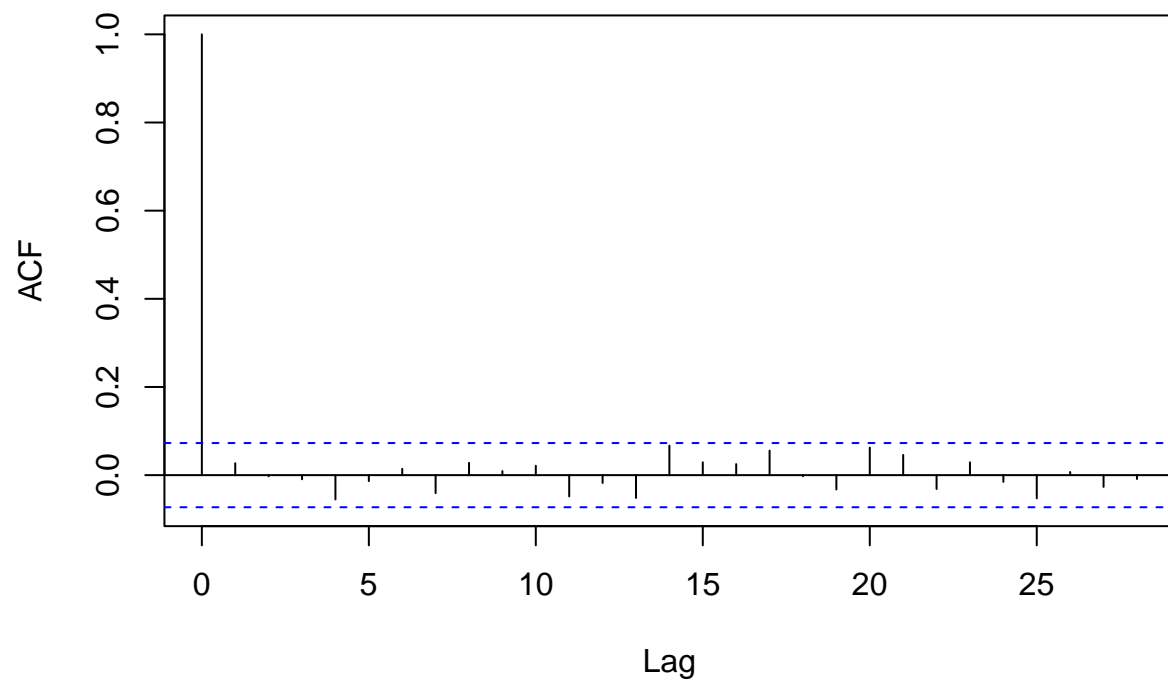


```
boxcox(result)
```



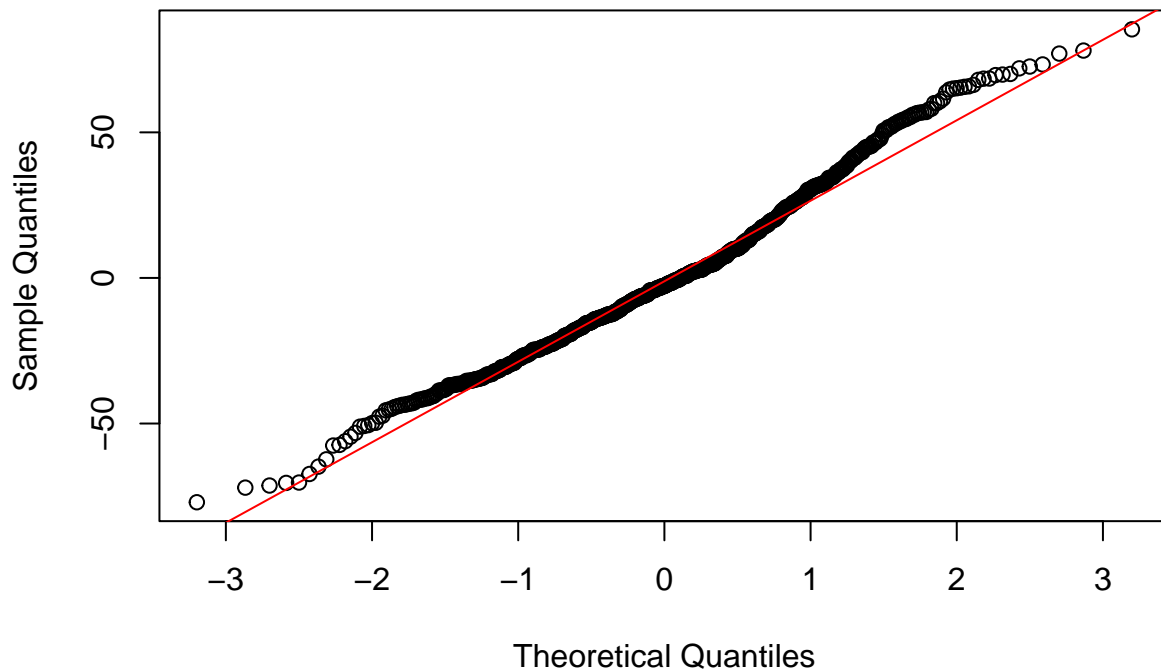
```
acf(result$residuals)
```

Series result\$residuals



```
qqnorm(result$residuals)
qqline(result$residuals, col = "red")
```

Normal Q-Q Plot



```
ystar <- log(diabetes2$Glucose)
diabetes2 <- data.frame(diabetes2, ystar)
```

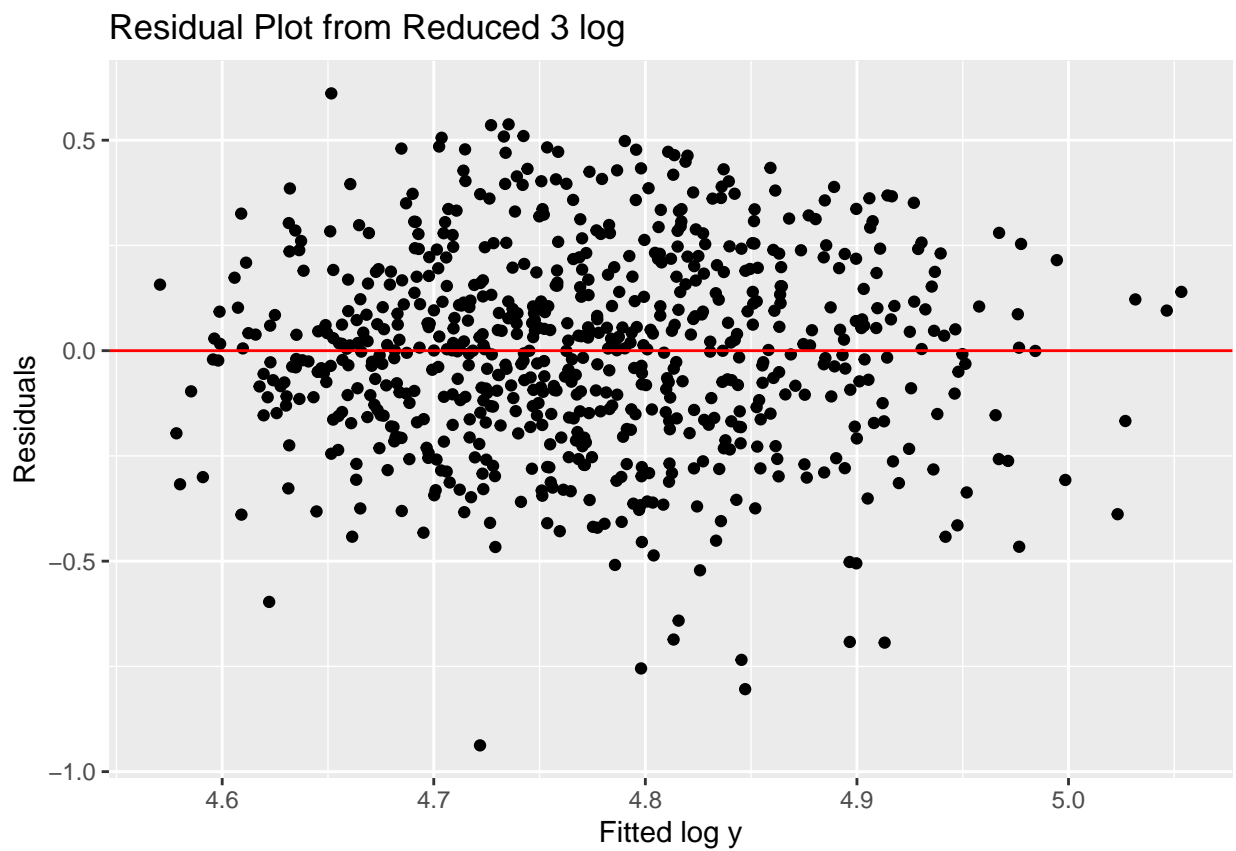
```
resultLog <- lm(ystar ~ BloodPressure + BMI + DiabetesPedigreeFunction + Age, data = diabetes2)
summary(resultLog)
```

```
##
## Call:
## lm(formula = ystar ~ BloodPressure + BMI + DiabetesPedigreeFunction +
##     Age, data = diabetes2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.93766 -0.15344  0.00361  0.15718  0.61112
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.2239516   0.0605403   69.771 < 2e-16 ***
## BloodPressure    0.0022203   0.0007865    2.823  0.00489 **
## BMI              0.0062395   0.0013532    4.611 4.74e-06 ***
## DiabetesPedigreeFunction 0.0714140  0.0267900    2.666  0.00786 **
## Age             0.0045093   0.0007921    5.693 1.82e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

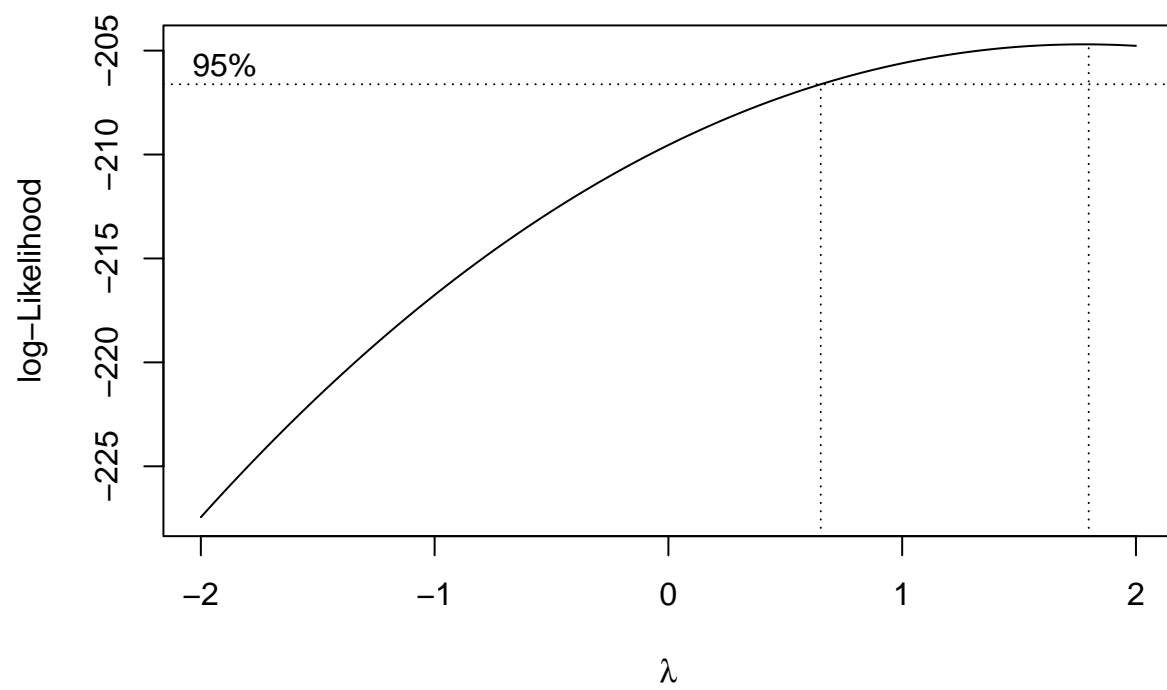
```
## Residual standard error: 0.2361 on 719 degrees of freedom
## Multiple R-squared:  0.1272, Adjusted R-squared:  0.1224
## F-statistic: 26.2 on 4 and 719 DF,  p-value: < 2.2e-16
```

```
yhatLog <- resultLog$fitted.values
resLog <- resultLog$residuals
diabetes2 <- data.frame(diabetes2, yhatLog, resLog)

ggplot(diabetes2, aes(x = yhatLog, y = resLog))+
  geom_point()+
  geom_hline(yintercept=0, color="red")+
  labs(x="Fitted log y",
       y="Residuals",
       title="Residual Plot from Reduced 3 log")
```

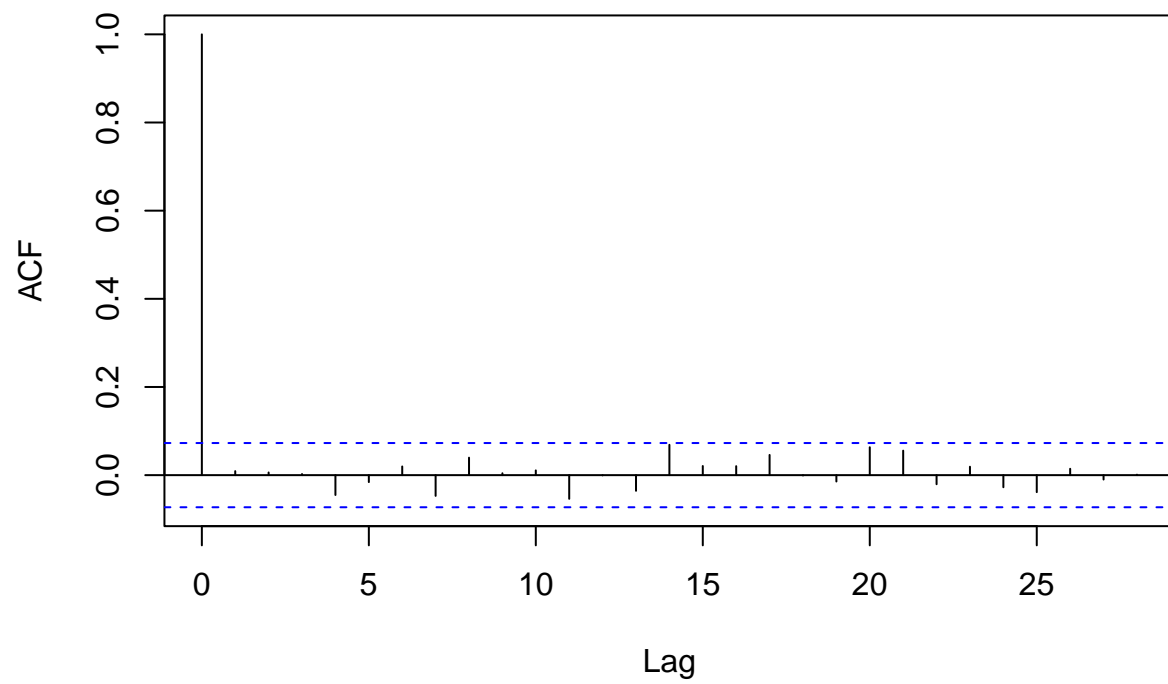


```
boxcox(resultLog)
```



```
acf(resultLog$residuals)
```


Series resultLog\$residuals



```
qqnorm(resultLog$residuals)
qqline(resultLog$residuals, col = "red")
```

Normal Q-Q Plot

