insurance

nehal linganur and zara nip

4/20/2023

```
insurance.df <- read.csv("insurance.csv")
```

```
fit.insur = lm(formula = charges ~ age + sex + bmi + children + smoker + region, data = insurance.df)
summary(fit.insur)
```

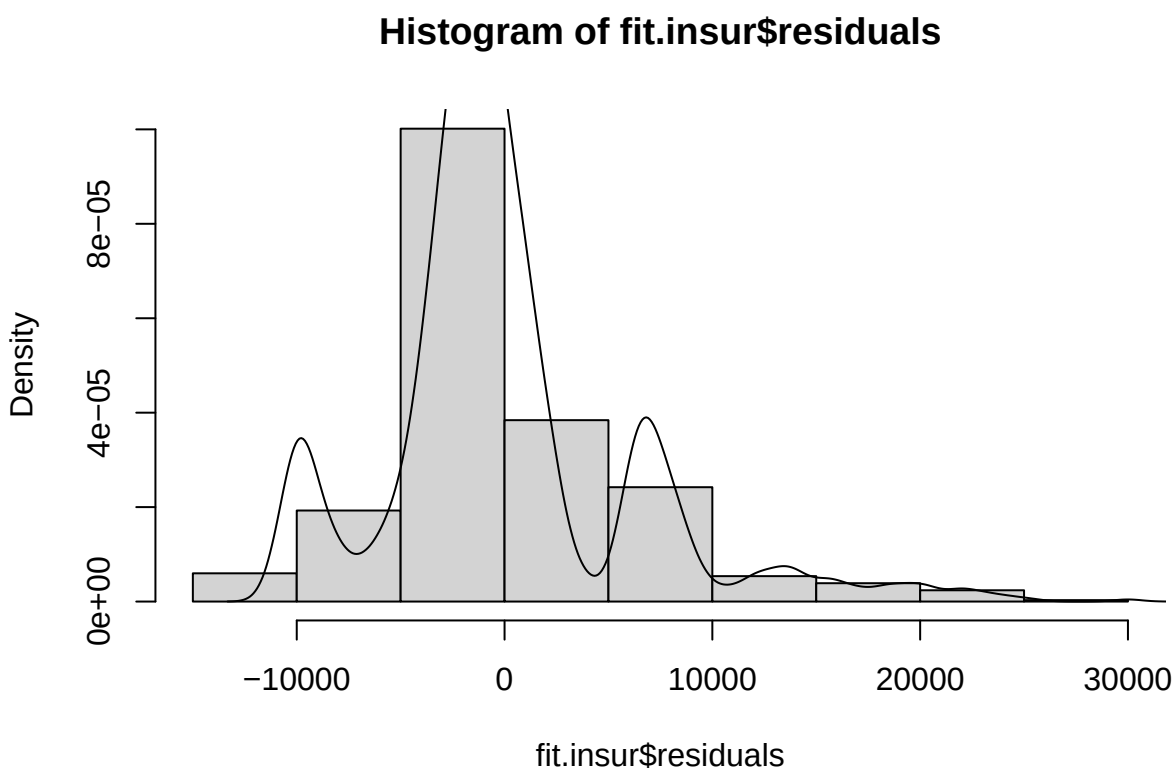
```
##
## Call:
## lm(formula = charges ~ age + sex + bmi + children + smoker +
##     region, data = insurance.df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11304.9  -2848.1   -982.1   1393.9  29992.8
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -11938.5      987.8 -12.086  < 2e-16 ***
## age                256.9       11.9  21.587  < 2e-16 ***
## sexmale           -131.3      332.9  -0.394 0.693348
## bmi                339.2       28.6  11.860  < 2e-16 ***
## children           475.5      137.8   3.451 0.000577 ***
## smokeryes        23848.5      413.1  57.723  < 2e-16 ***
## regionnorthwest   -353.0      476.3  -0.741 0.458769
## regionsoutheast  -1035.0      478.7  -2.162 0.030782 *
## regionsouthwest   -960.0      477.9  -2.009 0.044765 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6062 on 1329 degrees of freedom
## Multiple R-squared:  0.7509, Adjusted R-squared:  0.7494
## F-statistic: 500.8 on 8 and 1329 DF,  p-value: < 2.2e-16
```

```
fit.insur2 = lm(formula = charges ~ age + bmi + children + smoker, data = insurance.df)
summary(fit.insur2)
```

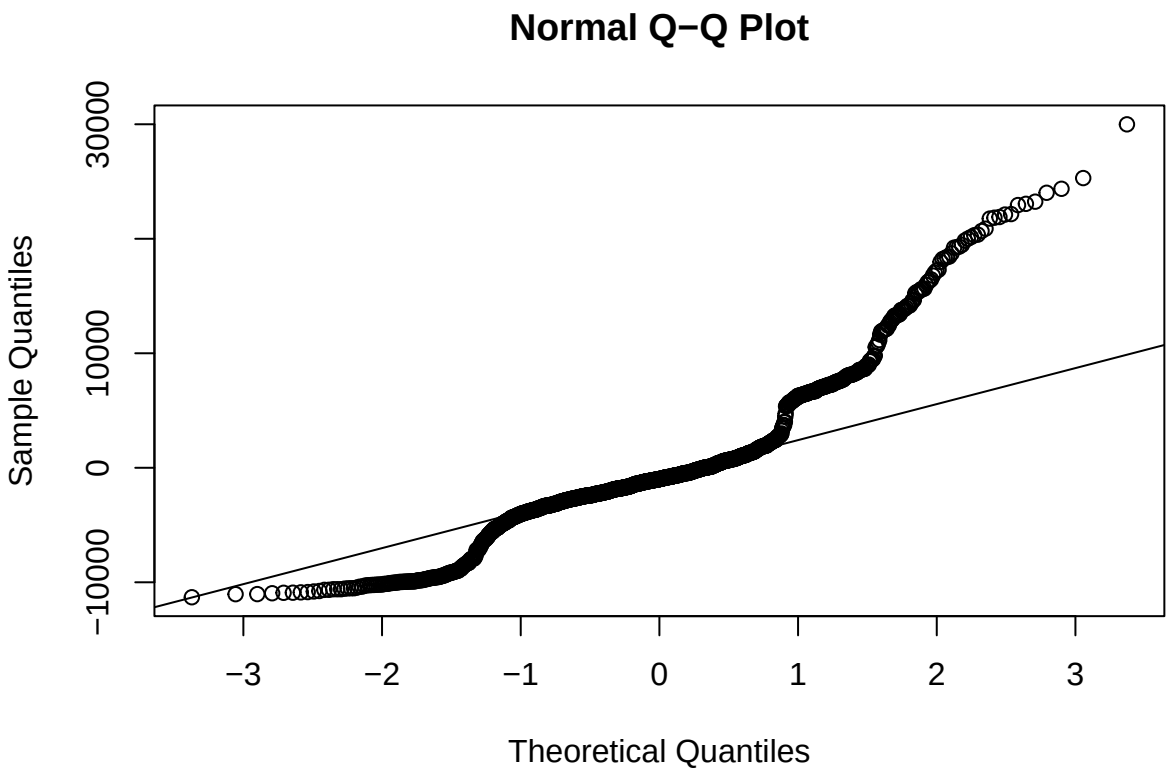
```
##
## Call:
## lm(formula = charges ~ age + bmi + children + smoker, data = insurance.df)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -11897.9  -2920.8   -986.6   1392.2  29509.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -12102.77     941.98 -12.848  < 2e-16 ***
## age            257.85      11.90  21.675  < 2e-16 ***
## bmi            321.85      27.38  11.756  < 2e-16 ***
## children       473.50     137.79   3.436 0.000608 ***
## smokeryes    23811.40     411.22  57.904  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6068 on 1333 degrees of freedom
## Multiple R-squared:  0.7497, Adjusted R-squared:  0.7489
## F-statistic: 998.1 on 4 and 1333 DF,  p-value: < 2.2e-16
```

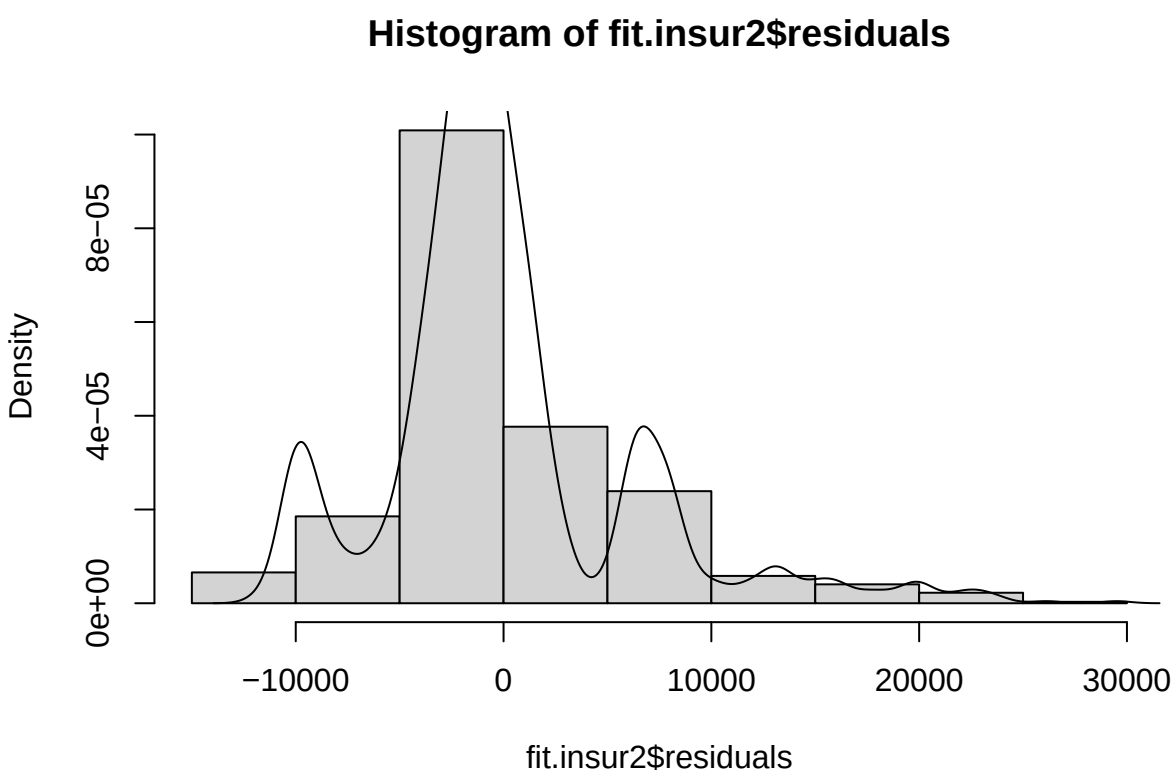
```
hist(fit.insur$residuals, prob = TRUE)
lines(density(fit.insur$residuals))
```



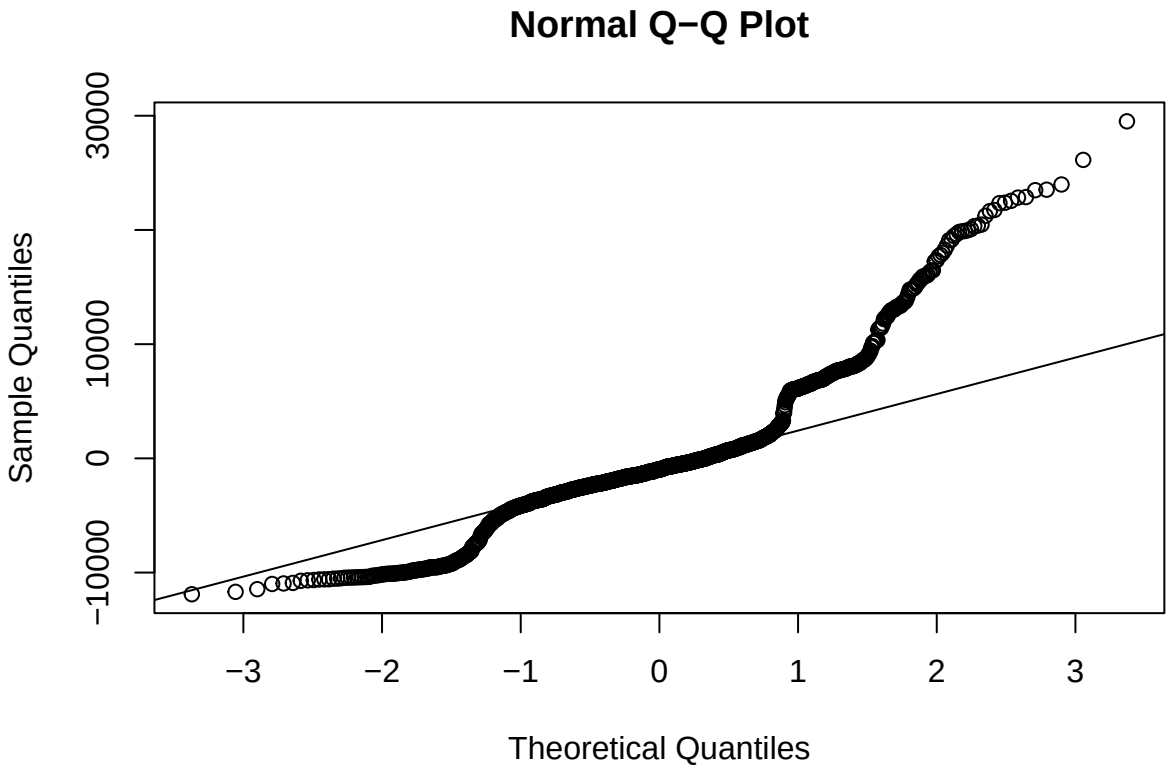

```
qqnorm(y=fit.insur$residuals)
qqline(y=fit.insur$residuals, datax = FALSE)
```




```
hist(fit.insur2$residuals, prob = TRUE)
lines(density(fit.insur2$residuals))
```




```
qqnorm(y=fit.insur2$residuals)
qqline(y=fit.insur2$residuals, datax = FALSE)
```




```
cont.table <- table(insurance.df$charges, insurance.df$bmi)
chisq.test(cont.table)
```

```
## Warning in chisq.test(cont.table): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  cont.table
## X-squared = 731886, df = 730792, p-value = 0.1827
```

Chi square test with bmi gives the lowest p-value