**STATS 500 hw3**

*Uniquname:changbai*

*Umid:92538445*

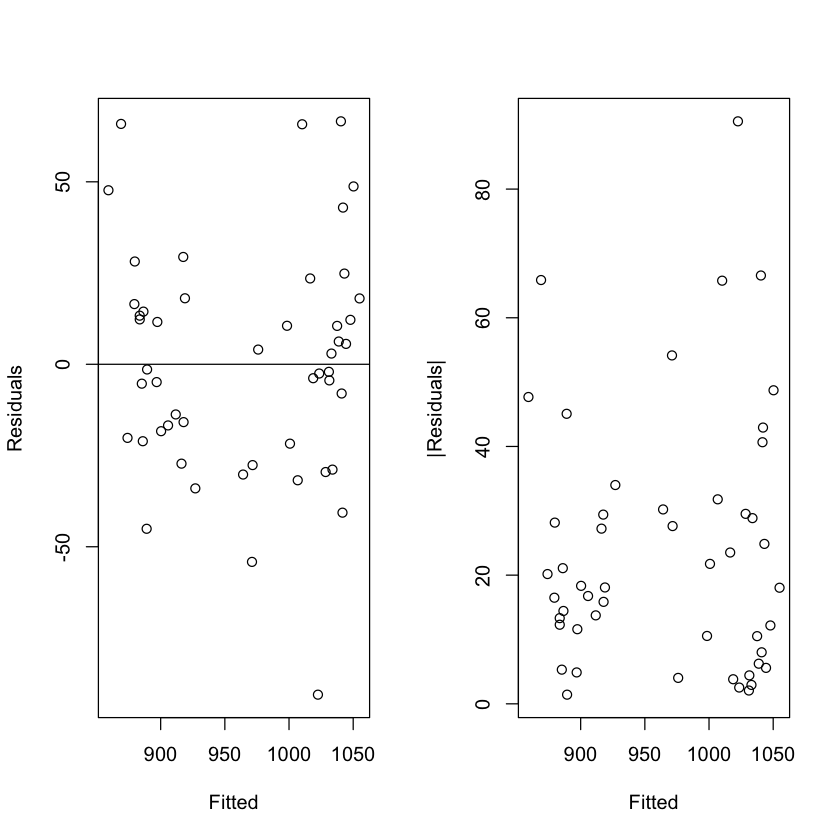
1. **Check the constant variance assumption for the errors and for evidence of non-linearity via residual plots, and adjust model as appropriate.**

par(mfrow = c(1,2))

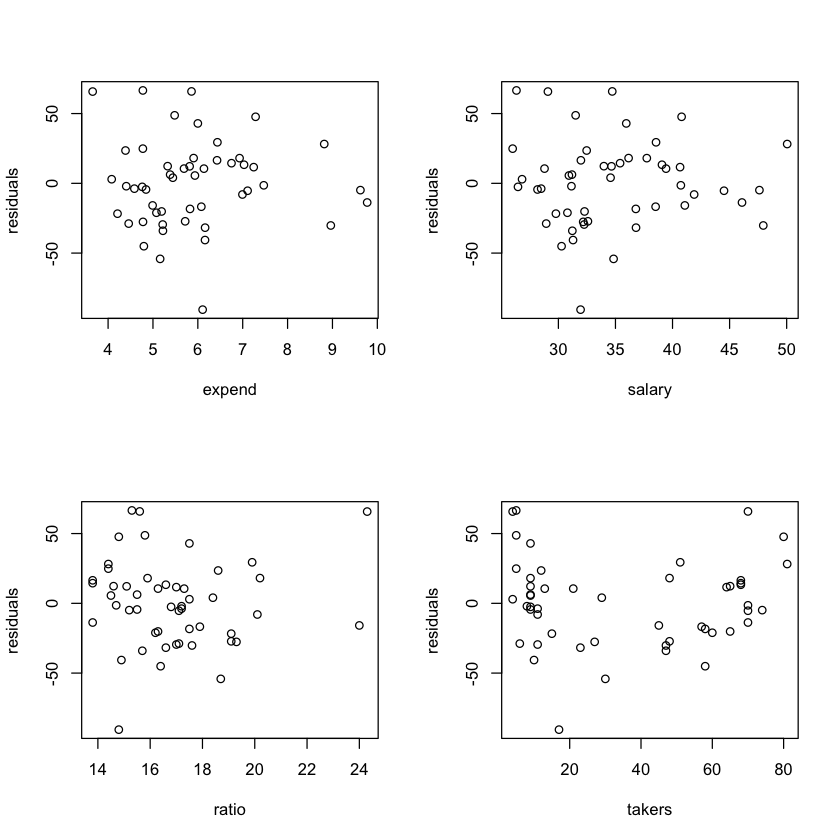
plot(fitted(sat.lm),residuals(sat.lm),xlab = "Fitted",ylab = "Residuals")

abline(h=0)

plot(fitted(sat.lm),abs(residuals(sat.lm)),xlab = "Fitted", ylab = "|Residuals|")



In the first plot, we see no evidence of non-constant variance.But, the second plot reveals non-linearity relationship. To check which predictor leads to this, we plot each predictor vs residuals.

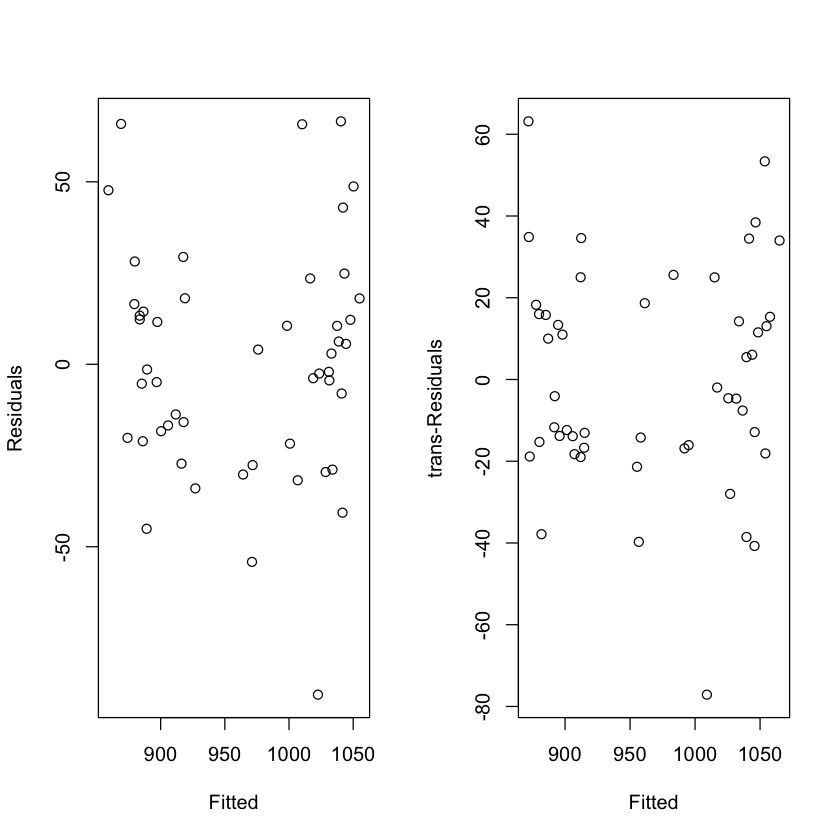


It seems the predictor(takers) has the non-linearity relationship with response.

So, we adjust the model like this:

*trans.lm <- lm(total ~ expend + salary + ratio + sqrt(takers), sat)*

Let’s see the difference between the old model and the new one.

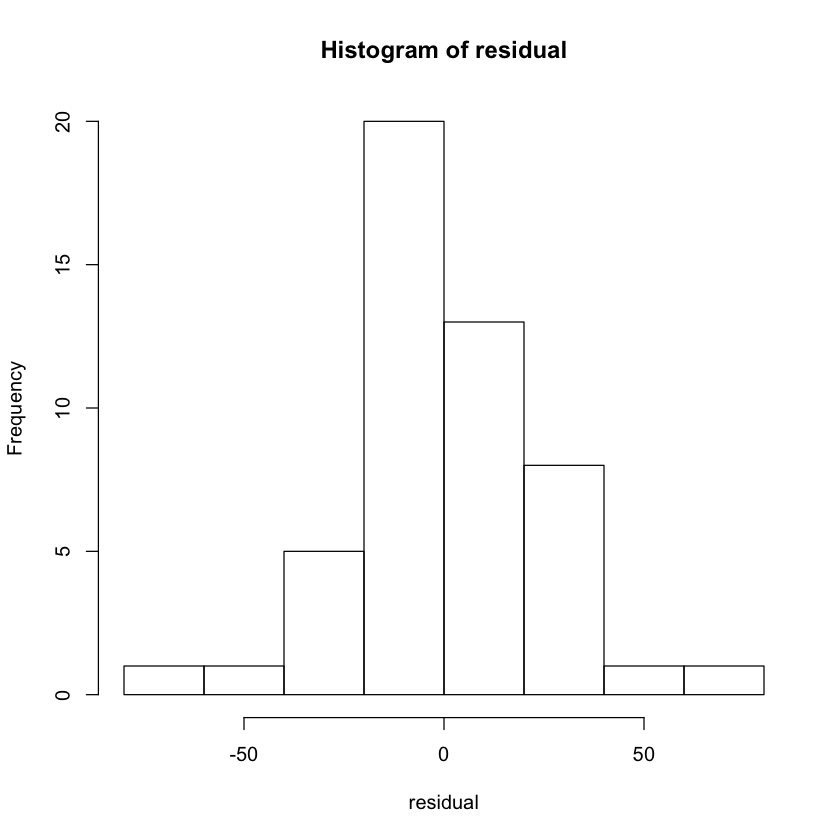


It seems better than before.

1. **Check the normality assumption.**

*residual = trans.lm$residuals*

*hist(sat.lm.resid)*

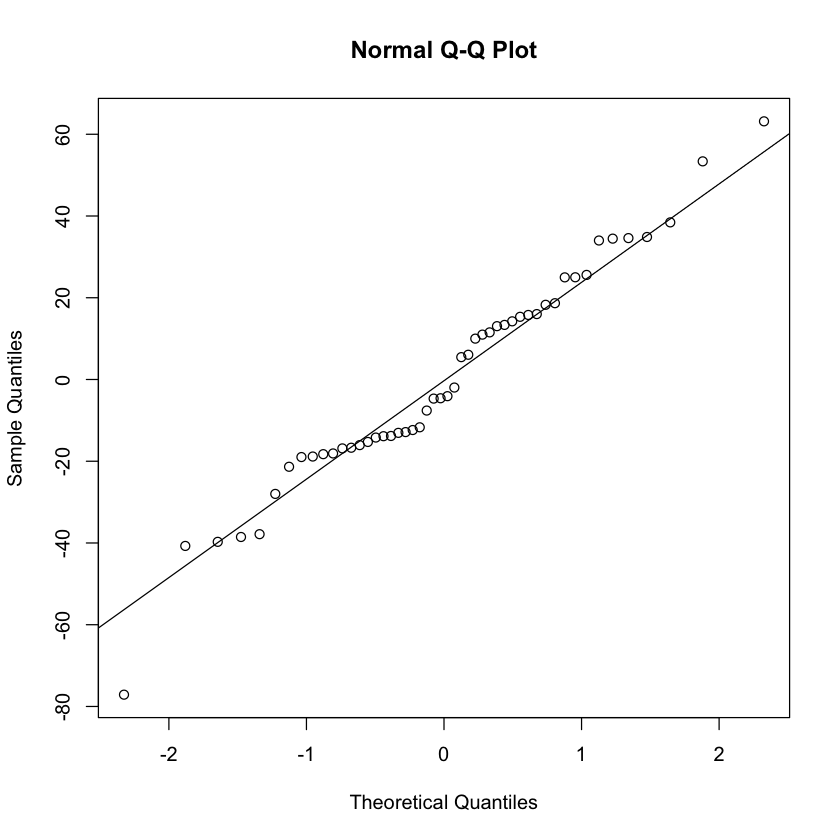


The histogram is expected bell shaped.

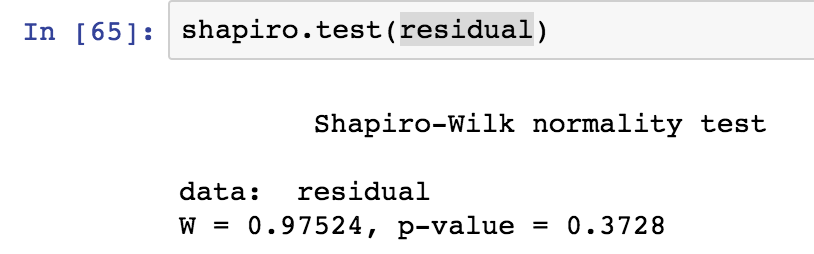
*#qqplot*

*qqnorm(residual)*

*qqline(residual)*

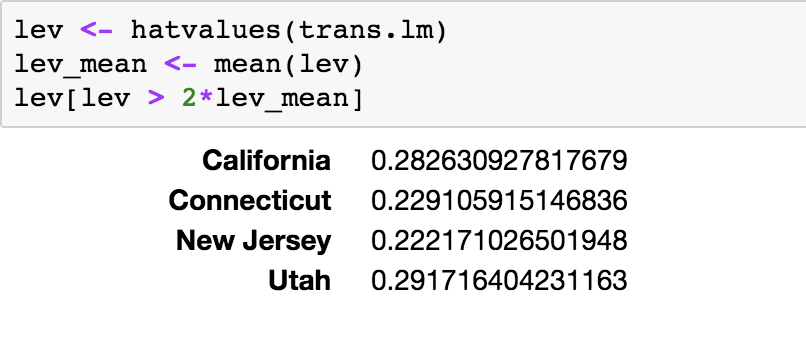


There are some outliers observed in the qqplot. Let’s check the Shapiro-Wilk test to see if this residual is normal or not.

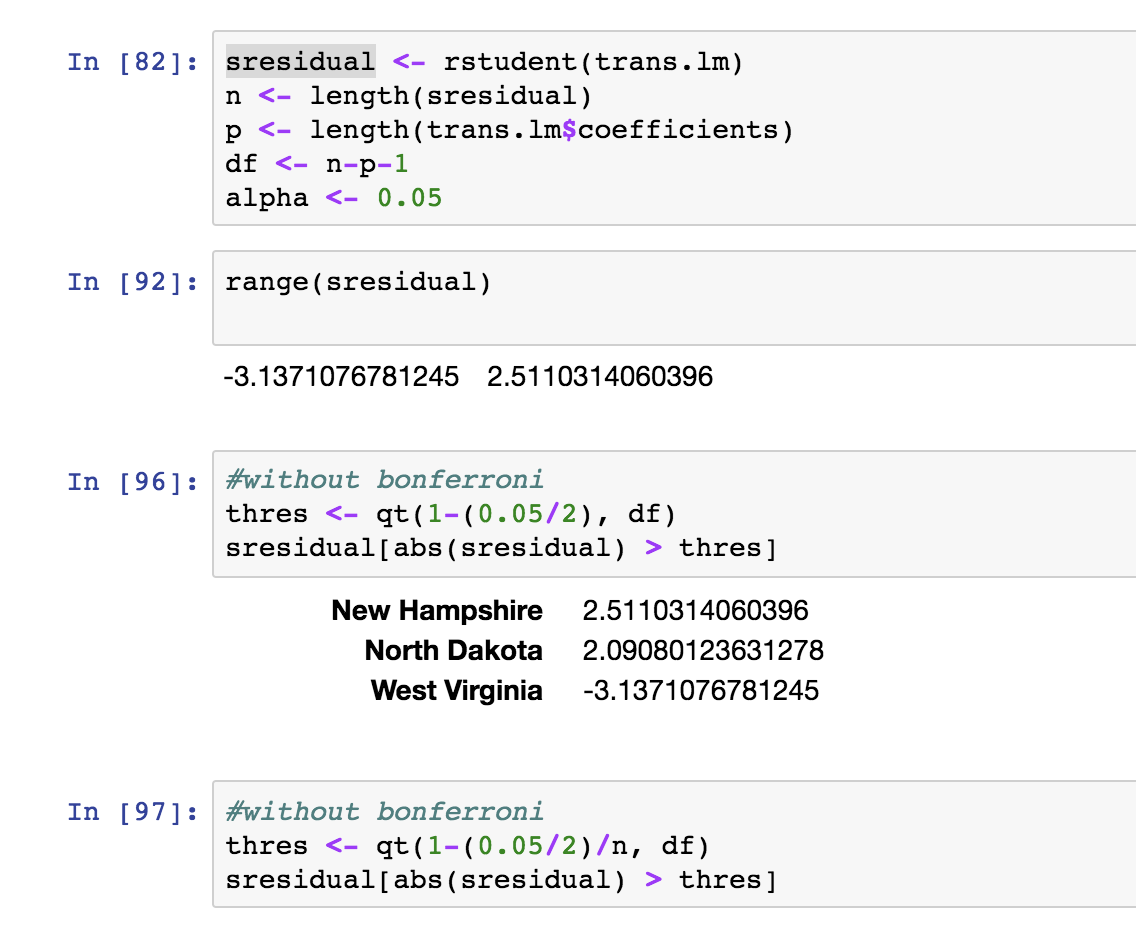


Since the p-value is small, and the qqplot seems to violate the normality assumption. We reject the assumption.

1. **Check for large leverage points.**



1. **Check for outliers.**



There are three outliers without Bonferroni.