• **Nonnormality**
  Can look at histograms and qqplots.

Comparing studentized residuals to $N(0,1)$:

```r
> qnorm(rstudent(lm.out),pch=16)
> qqline(rstudent(lm.out))
```

Using ‘qqPlot’ in the *car* library we can compare to $t_{n-k}$ distribution:

```r
> qqPlot(rstudent(lm.out), dist="t", df=45-2-2)
```

Some evidence of heavy-tails.

As mentioned before, near normality is probably fine for holding the $\alpha$ level error rate.

• **Nonconstant variance**
  Plot residuals against the fitted values:

```r
> plot(lm.out$fitted.values,lm.out$residuals)
```

The added variable plots can also give information on error variance by looking at a single covariate dimension (residuals from fitted line are $e_i$ values).
• **Nonlinearity**

When you look at the added variable plots, they should show a linear relationship.

![Added Variable plot](image)

Also, in your initial investigations, bivariate plots can also show nonlinearity.

• **Jointly influential points:**
  Check added variable plots for this.
  – Can be missed if only using delete-one diagnostics (COOKSD, DFFITS, ...).

• **Nonnormality:**
  Check qqplots of residuals (or $e_i^*$s).

• **Nonconstant variance:**
  Check residuals vs. fitted values (or use $e_i^*$s).
  – Added variable plots can also show this.

• **Nonlinearity:**
  Check added variable plots
  – Can also look at XY bivariate scatterplots.

• **Multicollinearity**
  Check VIFs
  – Also, try to notice unusual results.

**Summary of Diagnostics (parts 1 & 2)**

• **Independent variable (X-space) outliers:**
  Check for high leverage.
  – These have the potential to greatly influence fitted values.

• **Large regression residuals:**
  Check for large studentized residuals $e_i^*$.

• **Influence:**
  Many different quantities to check (COOKSD, DFBETAS, ...)
  – These can suggest points that have a big effect on the regression coefficients and fitted values.
  – We’ll use COOKSD, it brings together leverage and studentized residuals into a single measure of influence.

**Possible solutions, if problems arise**

• **Nonlinearity:**
  Transformation. Or add polynomial term for quadratic. Nonparametric method such as LOWESS fit.

• **Nonconstant variance:**
  Transformation. Or use weighted least squares if applicable (may discuss later).

• **Nonnormality:**
  Transformation. Or a nonparametric approach.

• **Outliers, or High Influence:**
  IF JUSTIFIED, remove. Could try ‘robust regression’. Fit model with and without point.

• **Multicollinearity:**
  No quick fix. Try variable selection or dimension reduction. Could try ‘ridge regression’.