**Class 10 Worksheet**

**Population: All diamonds for sale in Singapore**

**Sample size: n=48**

**Variables (type and roll):**

Size - quantitative, predictor, units are Carats

Price - quantitative, response, units are Singapore dollars

**Write out the SLR model for this data set:** Average Price = -259.63 + 3721.02\*Size

> diamonds\_SLR <- lm(Price ~ Carats, diamonds)

> summary(diamonds\_SLR)

Call:

lm(formula = Price ~ Carats, data = diamonds)

Residuals:

Min 1Q Median 3Q Max

-85.159 -21.448 -0.869 18.972 79.370

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -259.63 17.32 -14.99 <2e-16 \*\*\*

Carats 3721.02 81.79 45.50 <2e-16 \*\*\*

---

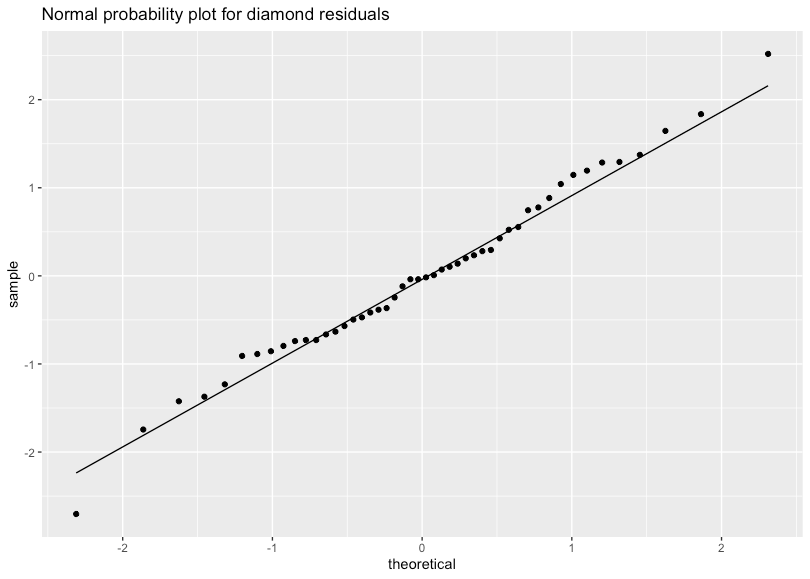
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31.84 on 46 degrees of freedom

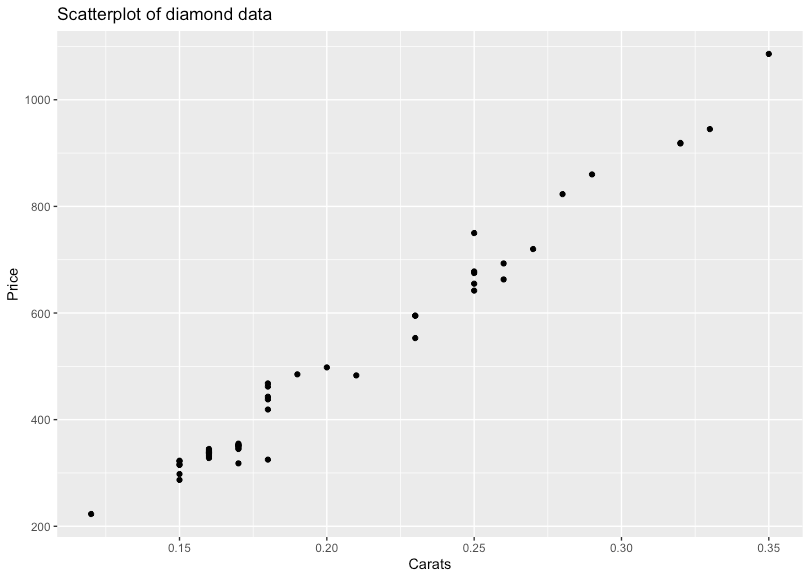
Multiple R-squared: 0.9783, Adjusted R-squared: 0.9778

F-statistic: 2070 on 1 and 46 DF, p-value: < 2.2e-16

**Normal Probability Plot for ~~Diamond Price~~** **Residuals:** Skip this plot for our Class 11 discussion.



**Scatter Plot for Carat and Price**



**Correlation Between Carat and Price: 0.989**

> cor(diamonds$Carats, diamonds$Price)

[1] 0.9890707

**Interpret the apparent relationship between these two variables:**

The correlation is really close to 1, this together with the scatter plot of the data indicates there is a strong linear relationship between diamond size and price.