STAT 400 - Discussion 4

Load in Wine Dataset

Add column names

```
library(readr)

url <- "https://archive.ics.uci.edu/ml/machine-learning-databases/wine/wine.data"

wine <- read_csv(url, col_names = FALSE)
```

```
Rows: 178 Columns: 14
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl (14): X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
colnames(wine) <- c(
  "Class",
  "Alcohol",
  "Malic_Acid",
  "Ash",
  "Alcalinity_of_Ash",
  "Magnesium",
  "Total_Phenols",
  "Flavanoids",
  "Nonflavanoid_Phenols",
  "Proanthocyanins",
  "Color_Intensity",
```

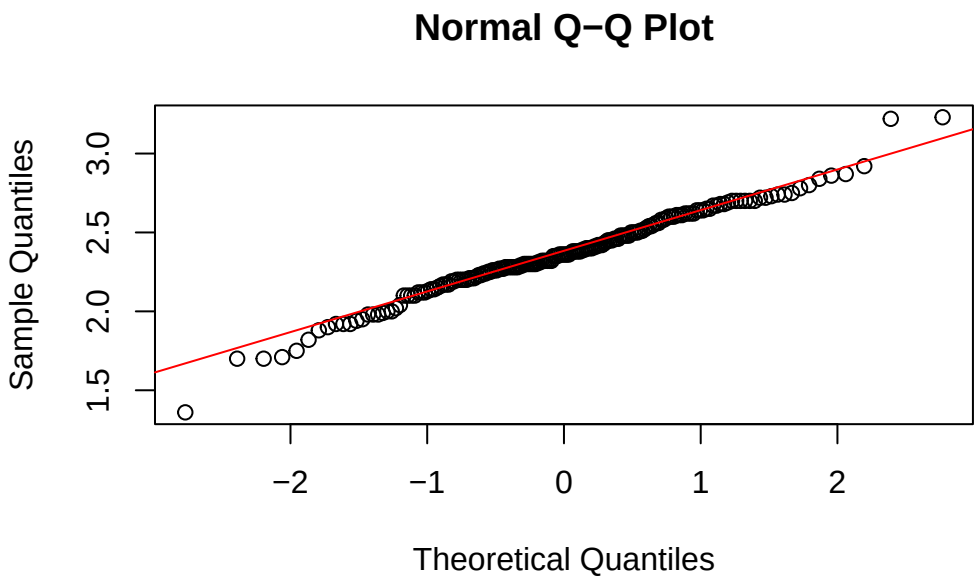
```
  "Hue",
  "OD280_OD315_of_Diluted_Wines",
  "Proline"
)

print(head(wine))
```

```
# A tibble: 6 x 14
  Class Alcohol Malic_Acid   Ash Alcalinity_of_Ash Magnesium Total_Phenols
  <dbl>   <dbl>      <dbl> <dbl>              <dbl>     <dbl>         <dbl>
1     1    14.2       1.71  2.43               15.6       127          2.8
2     1    13.2       1.78  2.14               11.2       100          2.65
3     1    13.2       2.36  2.67               18.6       101          2.8
4     1    14.4       1.95  2.5                16.8       113          3.85
5     1    13.2       2.59  2.87               21         118          2.8
6     1    14.2       1.76  2.45               15.2       112          3.27
# i 7 more variables: Flavanoids <dbl>, Nonflavanoid_Phenols <dbl>,
#   Proanthocyanins <dbl>, Color_Intensity <dbl>, Hue <dbl>,
#   OD280_OD315_of_Diluted_Wines <dbl>, Proline <dbl>
```

QQ Plot for Ash variable

```
qqnorm(wine$Ash)
qqline(wine$Ash, col='red')
```

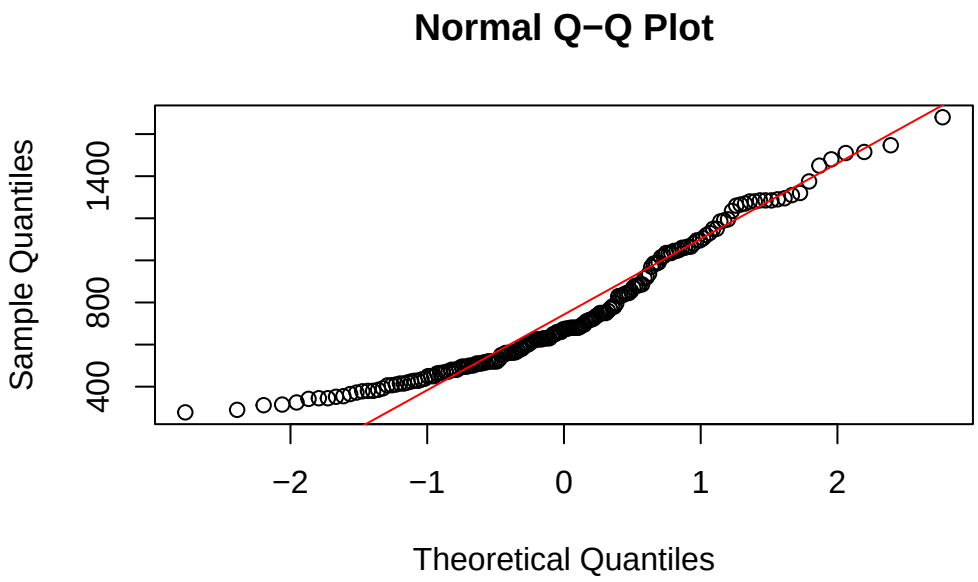


Assumption: Normal Distribution.

Since the majority of the points are lying along the expected distribution, we can conclude that the distribution of Ash values are normal. We can use parametric methods to asses with hypothesis testing.

QQ Plot for Proline variable

```
qqnorm(wine$Proline)
qqline(wine$Proline, col='red')
```



Assumption: Non-normal Distribution

Since the majority of the points are not lying along the expected distribution, we can conclude that the distribution of Ash values are not normal. If we were to analyze with various hypothesis tests, we would prefer to use nonparametric methods.