1 What is statistics:

Statistics is the science of collecting, organizing and interpreting the data.

1. The difference between Probability and Statistics
   (a) Population vs Sample
   (b) population mean and population variance and standard deviation.
      i. Measuring center: the mean ($\mu, \bar{x}$), median (M), mode
      ii. Measuring spread: the quartiles, the pth percentile, Q_1 and Q_3, the interquartile range (IQR)
      iii. The five-number summary and boxplots

      mean(x); var(x), median(x), quantile(x, 0.25),
   (c) Skew to right, Skew to left, symmetric ;
   (d) The (100p)th percentile of a distribution is often called the quantile of order p.

   q-q quantile-quantile plot , we expect the points $(y_r, \tilde{\pi}_r)$ lie close to a line through the origin with slop equal to 1. (ex 3.2-7, 3.5-4)

2. Displaying data with graphs : bar plot, pie chart, (ordered) stem-and-leaf plot, histograms,

3. Some basic R commands

bar and pie chart

> pie.sales <- c(0.12, 0.3, 0.26, 0.16, 0.04, 0.12) # or pie.sales=scan()
> names(pie.sales) <- c("Blueberry", "Cherry", "Apple", "Boston Cream", "Other", "Vanilla Cream")
> barplot(pie.sales)
> pie(pie.sales)
> stem(pie.sales)

histograms (ex 3.1 Mark McGwire's Home Runs)

```
> x=c(364,368,364,419,424,347,462,419,437,419,
+ 371,362,358,527,381,545,478,440,471,451,
+ 425,366,477,397,433,388,423,409,356,409,
+ 438,437,449,433,461,431,472,485,405,415,
+ 511,425,458,452,408,374,464,398,409,369,
+ 385,477,393,509,501,450,472,497,458,381,
+ 430,341,385,417,423,375,403,435,377,370)
> plot(x)
> mean(x)
> var(x)
> sqrt(var(x))
> summary(x)
```

Min. 1st Qu. Median Mean 3rd Qu. Max.
341.0 385.0 423.0 423.8 456.5 545.0
> fivenum(x)
[1] 341 385 423 458 545
> stem(x)
The decimal point is 1 digit(s) to the right of the |
34 | 1768
36 | 24468901457
38 | 11558378
40 | 35899957999
42 | 3345501335778
44 | 0901288
46 | 124122778
48 | 57
50 | 191
52 | 7
54 | 5
> par(mfrow=c(2,2))
> hist(x)
> hist(x, br=seq(337.5, 547.5, 30))
> hist(x, br=seq(337.5, 547.5, 30), prob=TRUE)
> hist(x, br=20)
> par(mfrow=c(1,1))
> hist(x, br=seq(337.5, 547.5, 30), prob=TRUE)
> lines(density(x), col="red") # a density estimate of x
> rug(x)

q-q plot

> qqnorm(x) #quantiles of x with respect to the values of expected under a normal law
> qqline(x, col=2) #draw a line with slop=1
> qqplot(y, x); abline(0,1, col=2) # quantiles of x w.r.t. the quantiles of y

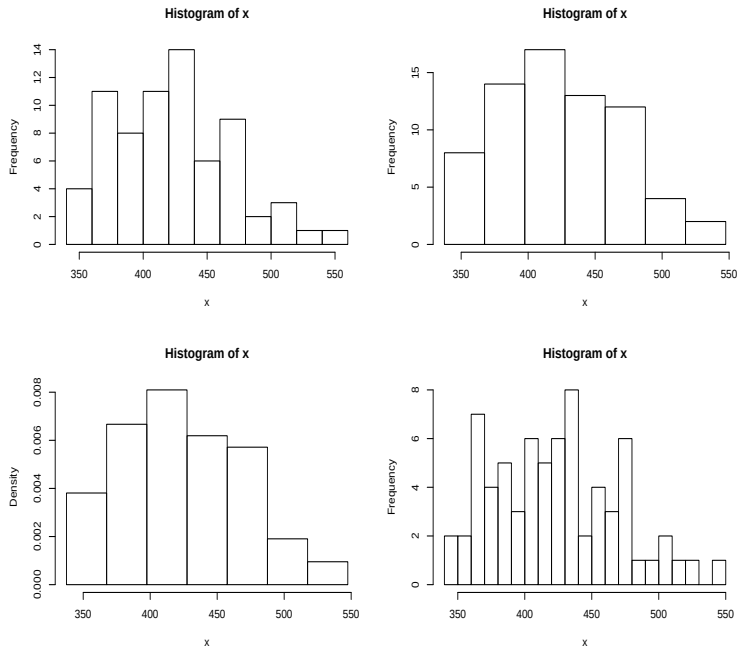

Figure 1: Example for histogram

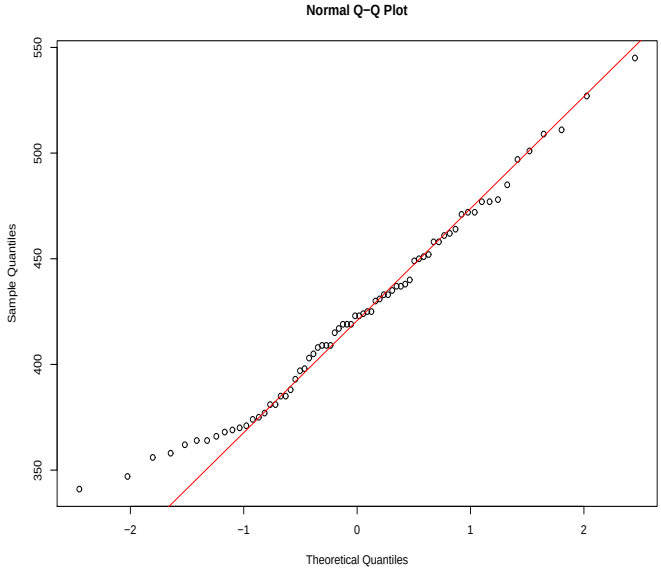

Figure 2: q-q normal plot