

Stem and leaf plots

5 marks

Stem and leaf plots. Install the package `aplpack`; the function `stem.leaf(...)` from that package will be used to construct the displays. You will work with the following data:

```
library("aplpack")
```

```
## Warning: package 'aplpack' was built under R version 4.0.3
```

```
help("stem.leaf")
```

```
## starting httpd help server ... done
```

```
set.seed(1234567) # Use this so we all get the same values
x30 <- rnorm(30)
x100 <- rnorm(100)
x500 <- rnorm(500)
x1000 <- rnorm(1000)
x1000 <- rnorm(10000)
```

Stem and leaf plots are then generated as follows:

```
library(aplpack)
#stem.leaf(x10000)
```

- a. **(3 marks)** What characteristics of the data can you see (or easily determine) from the stem and leaf display? DO NOT hand these displays in!

We see a histogram with x,y axis reversed and some additional information.

Specifically, we see the number of data in each bin in terms of bin height along the x axis.

We see number of one sided data that are as extreme as or more extreme than data in each bin in terms of numbers.

We see each data with precision of leaf unit in terms of numbers.

we see the shape of distribution density, in this case, we see bell shape curve and symmetric around $x=0$

we also see extreme points not shown in the histogram written as high precision float at bottom of histogram.

- b. **(2 marks)** How does the stem and leaf adapt to increasing sample size? Suppose a sample of $n = 10,000$ were to be displayed on a single sheet of paper (using the default values for `stem.leaf()`). What would be the problems, if any, with that?

If the smallest bin width is still 0.2, there will be too much data in one bin that one line could not display all of them.

Each number must at least take some nontrivial area in our display. So the total area of these numbers may be larger than area of the paper.

The number of one sided data that are as extreme as/more extreme than our current bin is getting too large to hold intuitive meaning to us.

There will be too many outliers displayed in high precision number, we do not have a good visualization on how extreme they are or their distributions.