Lecture 9 (Ch 2-3)

One more thing in Ch 2.

This business of estimating pop. parameters refers to any parameter. Specifically, $\bar{x}$ and $s$ provide point estimates for $\mu$ and $\sigma$, respectively, of the normal distr. if the data come from a normal distr. to begin with.

But how do we know if our data come from a normal?

Easier Q: how do we know if our data come from std. normal?

If the histogram is consistent with a std. normal, then the quantiles/percentiles of data should be equal/comparable to those of the distr. To see that, plot "qq plot".

In R: `qqnorm(x)`

If the data are not from std. normal, but from $N(\mu, \sigma)$, the only thing that changes is that the slope becomes 0, and the intercept becomes $\mu$. Not too obvious, but plot in book.
from histogram, it's hard to tell if precip is normal, especially because its shape depends on bin size.

Histogram of precip

The plot looks linear, mostly! So, data are consistent with a normal.

In fact, it looks like 2 different normals with diff μ's, same σ (slope). i.e. bimodal.
Thus far, our focus has been on 1 column of data, and 1 variable. I.e. univariate analysis.

With 2 (or more) variables, we can do all of the above, but we can also ask about the relationship between them.

For continuous data: scatterplot

categorical data, later

<table>
<thead>
<tr>
<th>Data</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

Although one purpose of a scatterplot is to summarize and display the relationship between 2 cont. variables, there is nothing that can fully substitute it.

I.e. Given data on 2 vars., do the scatterplot!

Of course, histograms each one, too.

are not unusual. In fact, they are common, (and even necessary, later).
A scatterplot is the best device for displaying and studying the relationship between data on 2 continuous variables or association.
hw-D: Do a q-q plot of each of the 2 cont. vars. in your data collected in hw-A. (By computer)
Describe the result.

Note: If you find out that there is not much you can say about the q-q plot, it may be that your data is not appropriate. This will be another chance to correct the error, because later you will be doing more hw problems using your data.
So, see me, if you are not sure.

hw-P: Make a scatterplot of the 2 cont. vars collected in hw-A. Describe the relationship.
(By R or by hand).
Same comment as above.