Statistical Inference

The statistical world

- Consists of:
  - objects or people of interest to us (*individuals*)
  - things measured or counted on those individuals (*variables*)
- About the individuals:
  - which ones do we care about? All of them, the *population*.
  - which ones do we know about? The ones we happened to look at, the *sample*.
- Sample is (or should be) randomly chosen from population, with no favoritism.

Sample to population: confidence interval

- Want to know about population (parameter), but don't. Only have sample (statistic). Eg. population mean, only have sample mean.
- Logic:
  - *If* we knew about population, could figure out kinds of samples that might appear (math).
  - In particular, can figure how far apart sample statistic and population parameter might be.
  - Use this to construct *confidence interval* for population parameter: says eg. "based on my sample, I think population mean between a and b".

Test of significance

- Or:
  - might have theory leading to *null hypothesis* (eg. population mean is 20) and *alternative hypothesis* (eg. population mean not 20).
  - This leads to *test of significance* (hypothesis test): “based on my sample, I think pop. mean is (is not) 20”
  - Done by choosing α (eg. 0.05), calculating *test statistic* and *P-value*. If P-value $< \alpha$, *reject null*: have evidence in favour of alternative.
- Math producing inference procedures can be difficult, but calculations (with software) and interpretations need not be.