

# Stat 204 Sampling

Prof. Art Owen

Monday & Wednesday 1:30–2:50, room TBA

Where does data come from? Often it is a byproduct of some other process such as a billing system that logs transactions. Every once in a while it comes from a purposeful experiment. (Hurray for stat 263/363!) The other place is sampling. Somebody sets out to purposely gather the needed data. They sample from some population of people or events or items. There is then a tradeoff between accuracy and cost, and clever tactics can help you get the best outcome.

The classic use case for sampling is surveying for opinions. That subject is in a severe crisis right now. Response rates have fallen below 10%. There is now a lot of interest in methods to mitigate non-response bias, including weighting of survey data, and even moving away from probability sampling.

## Classical topics

- Simple random sampling,  $n$  of out  $N$  items.
- Unequal probability sampling (Horvitz-Thompson).
- Raking and other weighting schemes.
- Ratio and regression estimators.
- Stratified sampling.
- Cluster sampling.
- Systematic sampling.
- Two-stage sampling.

## Less traditional topics

- Internet surveys.
- Network sampling and link tracing for hard to reach populations.
- Reservoir sampling (from a stream of data).
- Capture-recapture.
- Small area estimation.
- Transect sampling.