Population $X \sim ? \quad \Rightarrow \quad X_1, \ldots, X_n \quad \text{Sample}$

Parameter $\theta \quad \leftarrow \quad \hat{\theta} \quad \text{Statistic}$

1. Confidence intervals for $\mu$.

   - $X$ normal?
     - YES $t$ is exact
     - NO $n$ large

   - $n$ large
     - YES $t$ is robust
     - NO “proceed with caution”

2. What if the underlying distribution is not known and/or the sampling distribution of the statistic is not well understood?

3. The bootstrap. Use the sample to model the population.
   - (a) Act as if $\frac{1}{n}$ the population is equal to each element of the sample.
   - (b) Choose many (thousands) of samples of size $n$ from this population! (This is equivalent to choosing samples of size $n$ from the sample with replacement.)
   - (c) Compute properties of the statistic by using the simulated samples to generate an approximation of the sampling distribution of the statistic.

Homework - due Monday, April 13, 2009

1. Read Section 5.7.

2. Refer to the dataset of useless data about students in our class collected on the first day of class. Pretend our class is a random sample of all Calvin students. [http://www.calvin.edu/~stob/data/uselessdata.csv](http://www.calvin.edu/~stob/data/uselessdata.csv)
   - (a) Use the bootstrap to construct a 95% confidence interval for the average amount of cash carried by a Calvin student to class on the first day of classes.
   - (b) Compare the result in (a) to the 95% confidence interval generated by using the standard (based on the $t$-distribution) procedure. (A comparison could involve its location and its size.)
   - (c) Use the bootstrap to construct a 95% confidence interval for the standard deviation of the amount of cash carried by a student to class.
Useful R

```r
> aircondit$hours
[1]  3  5  7 18 43 85 91 98 100 130 230 487
> r=replicate( 10000,  mean( sample(aircondit$hours,12,replace=T)))
> quantile(r,c(.025,.975))
    2.5%    97.5%
   46.750   191.002
> t.test(aircondit$hours)

   One Sample t-test

data:  aircondit$hours
  t = 2.7483, df = 11, p-value = 0.01895
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 21.52561 194.64105
sample estimates:
  mean of x
    108.0833
```