1 Lecture2 Friday 01/19/01

Homework 1:
chap. 7, exs : 1, 8, 19, 22, 45, 50, due next wednesday in class.

Sectionning:
labs....start next week

```
Monday 2:15: (15 people)
 nmfong, ndaviden, kpilner, rona, ytai, pbarrile, apiruk,
 salvekar, sryoon, bshaby, vsandin, emily, shunta, drubin, hlwang

Tuesday 6:30pm: (14 people)
 jesseb, fusionic, jgifford, michali, wptang, genewong, bidemi, shubbard,
 yzou, jjmkeever, subarna, keljr, opium, xyoli

Wednesday 2:15pm: (14 people)
 keylee, nbastian, hugh2, richardp, justlee, ncwang, hwu591, dhorner,
 kaleeg, leahob, nzhang, kcorby, shairhee, hsuak

Unassigned to these dates: adho
```

Theorem 1 *With simple random sampling $E(\bar{X}) = \mu$.*

See proof page 191, it is based on the fact that $E(X_i) = \mu$.
Example:
Special case, dichotomous variables $\xi_1 = 0, \xi_2 = 1$, $p = n_2/N$ $\bar{x} = \hat{p}$

Definition 1 *An estimate is unbiased if its expectation equals the parameter we want it to estimate.*

Next we are going to look at the variance of the mean:

$$var(\sum b_i X_i) = \sum_i \sum_j b_i b_j cov(X_i, X_j)$$

so

$$var(\bar{X}) = \frac{1}{n^2} \sum \sum cov(X_i, X_j)$$

If the sampling were with replacement, we would have: X_i and X_j independent and thus their covariances are zero, for $i \neq j$ and so on.

$$var(\bar{X}) = \frac{1}{n}\sigma^2$$

or the $se(mean) = \frac{\sigma}{\sqrt{(n)}}$

As it is they aren't independent, although if n is small compared to N, the dependence can be very weak.

So we need to compute $Cov(X_i, X_j)$

Lemma

$$Cov(X_i, X_j) = -\frac{\sigma^2}{N-1}$$

Proof: (SEE THE BOOK, page 193, I did the details in class).

Notice that this variance is similar to the variance in the sampling with replacement case which could have been just σ^2/n, and in particular there is a pseudo-square root law for the standard error of the mean, if you want to double the precision (ie here variance of the estimator, you have to quadruple sample size).

Th other factor in the precision is of course the population variance, of course for σ small, little variation in the population values and the estimator will not vary much.

Corollary:

$$Var(\bar{X}) = \frac{\sigma^2}{n}(\frac{N-n}{N-1}) = \frac{\sigma^2}{n}(1 - \frac{n-1}{N-1})$$

Comment:

The last factor will be quite close to 1 when the sample is nearly as big as the population size and the estimate will be much less uncertain.