MATH 232: Engineering Mathematics
Reading Guide for LAS, Section 4.6: Sampling

Goals: 1. To learn of some sampling techniques, why they get used, and what are the pitfalls (if any).

2. To learn useful R commands for simulating the taking of random samples from data.

3. To begin to learn to identify sample from population (correspondingly statistic from parameter).

Read: Section 4.6 of LAS

Useful R commands:

<table>
<thead>
<tr>
<th>Commands</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample()</td>
<td>take a sample from a vector of data</td>
</tr>
<tr>
<td>replicate()</td>
<td>repeat some process a set number of times</td>
</tr>
</tbody>
</table>

Examples (without output):

```r
> sen = read.csv('http://www.calvin.edu/~scofield/data/tab/rc/senate99.dat',
    sep='\t', header=T)
> names(sen)
> sen$sex == 'm'
> sum(sen$sex == 'm')
> sum(sen$party == 'Dem')
> sample(1:100, 10)
> sen[sample(1:100, 10), ]
> sen[sample(1:100, 10), ]$party == 'Dem'
> sum(sen[sample(1:100, 10), ]$party == 'Dem')
> x = replicate(10000, sum(sen[sample(1:100, 10), ]$party == 'Dem'))
> histogram(x, breaks=0:10)
```

Terms to know:

- population vs. sample, parameter vs. statistic, census, inference, convenience sample, simple random sample, stratified random sample, sampling error, oversampling

Questions you should be able to answer:

1. Some named methods for taking samples include convenience sampling, simple random sampling, stratified sampling, voluntary response sampling (like that used by
Ann Landers in her survey of “whether parents would have children if they had it to do all over again”). Which of these, in general, produce useful samples? Are the others guaranteed to result in more sampling error?

2. Suppose phone numbers are randomly selected from around the country, and an interviewer is assigned to call some of these numbers.

   (a) If there are a number of phone numbers dialed at which the person who picks up refuses to participate, why might this tend to damage survey results?

   (b) Just because a person answering the phone agrees to participate does not mean other things cannot go wrong with the interview. Describe a couple.

3. When we took simple random samples of 1999 Senators in class (see example commands above), we generally wound up with a sample that reflected that Senate’s breakdown of sex and party affiliation. Could one or both of these variables be used to divide the population of Senators into strata? If so, would there be some advantage in doing so?

4. Suppose a certain college had 300 faculty members, 75 of whom were female. One way to select a faculty committee of 8 members would be to require 2 of them to be female, selecting those two at random, and then filling out the committee with 6 randomly-selected males. Would a committee built this way be considered an SRS? Explain.