Professor Notes About the "Sampling Distributions" Homework

- XXX

Simulating a Sampling Distribution

1. Results for the samples are in Table 1.

   Table 1. All possible samples of n=2 from the numbers 2 through 9. Each column represents the two values in a possible sample.

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
[1,]    2    2    2    2    2    2    2    3    3     3     3     3     3     4
[2,]    3    4    5    6    7    8    9    4    5     6     7     8     9     5
     [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25] [,26] [,27]
[1,]     4     4     4     4     5     5     5     5     6     6     6     7     7
[2,]     6     7     8     9     6     7     8     9     7     8     9     8     9
     [,28]
[1,]     8
[2,]     9
```

2. Results for the means are in Table 2.

   Table 2. All possible means from samples of n=2 from the numbers 2 through 9.

```
 [1] 2.5 3.0 3.5 4.0 4.5 5.0 5.5 3.5 4.0 4.5 5.0 5.5 6.0 4.5 5.0 5.5 6.0 6.5 5.5 6.0
[21] 6.5 7.0 6.5 7.0 7.5 7.5 8.0 8.5
```

3. The histogram of means is shown in Figure 1. The distribution is perfectly symmetric.

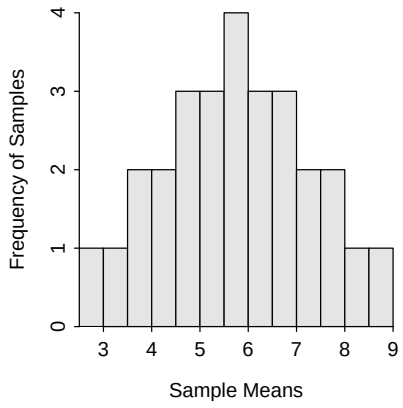

   Figure 1. Histogram of all means from samples of n=2 from the numbers 2 through 9.

4. Both the mean of the 28 sample means and the mean of the original set of numbers (the population) are 5.5. This observation is a demonstration of the definition of an unbiased statistic.
5. The standard deviation of the 28 sample means is 1.53. This standard deviation is called a standard error. The standard error would be smaller if samples of n=3 had been taken rather than n=2.

Precision and Accuracy

1. The numbers 59, 60, 60, 61 represent values that are accurate and precise (i.e., values are close together and centered on (average out to be) 60).
2. The numbers 45, 55, 65, 70 represent values that are accurate but imprecise (i.e., values are far apart and centered on 60).
3. The numbers 69, 70, 71, 72 represent values that are inaccurate but precise (i.e., values are close together and NOT centered on 60).
4. The numbers 75, 85, 95, 105 represent values that are inaccurate and imprecise (i.e., values are far apart and NOT centered on 60).

R Appendix.

```
library(NCStats)
( smpls2 <- combn(2:9,2) )
( mns2 <- as.numeric(combn(2:9,2,mean)) )
mean(mns2)
mean(2:9)
sd(mns2)
hist(~mns2,w=0.5,xlab="Sample Means",ylab="Frequency of Samples")
```