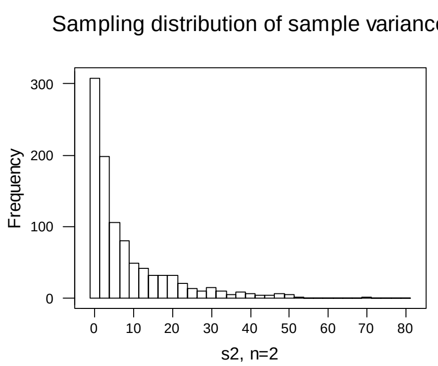
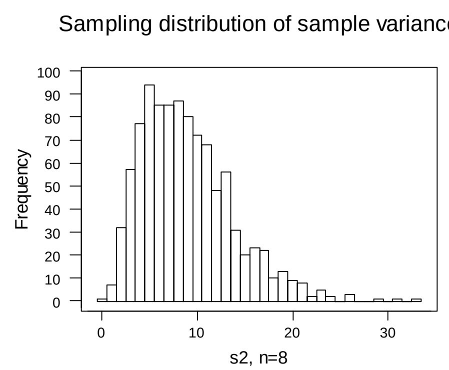
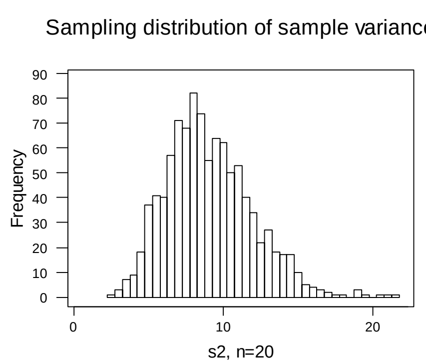
STA362: Introduction to Statistics
Illustration of the sampling distribution of s^2

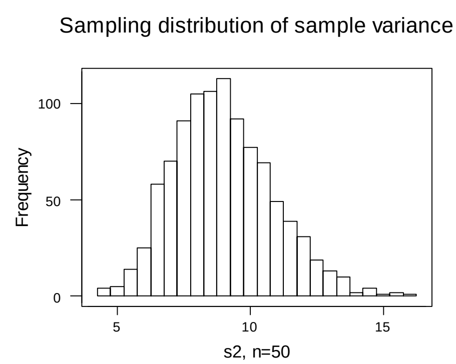
Since we often will be using sample data to infer about a population mean μ , we might also have to use the sample standard deviation s as a estimator of the population standard deviation σ.   Thus, it is useful to know about the sampling distribution for this measure of variability. Unfortunately, s is a statistic that doesn't behave like $\overline{x}$ from sample to sample.   Distributional theory dictates that we study the sampling distribution of s^2 (not s), and that we **assume the population from which sampling takes place is normal to begin with** (not a CLT assumption).

Here are four 1000-rep simulated sampling distributions of s^2 from a normal population with $\mu = 10$ and $\sigma = 3$ (i.e., $\sigma^2 = 9$).   We observe the behavior of s^2 under a variety of sample sizes: $n = 2$, $n = 8$, $n = 20$, and $n = 50$.









Note that the shape of this distribution always remains positively skewed, even for large n. This will always be true of the sampling distribution of the sample variance s^2.   Also, the severity of the skew lessens as the sample size increases, so a good distributional model for this statistic must reflect this. In class, we will develop the **chi-square (χ^2) distribution** for modeling the sampling distribution of the sample variance s^2.

We note here (without simulation proof) that the resulting sampling distributions for s^2 would differ dramatically from the above illustrated behavior were we sampling from a distinctly non-normal population. The normality assumption for the originating population is crucial to ensure the accuracy of the χ^2 model.