STAT 301 HW6 due Mon Oct 5

You may work with one partner on this assignment, submitting one report with both names, provided that both students contribute substantially to the work. Word-processed reports are preferred to hand-written ones. Please copy/paste relevant computer output into your report as appropriate.

Sampling Smokers?

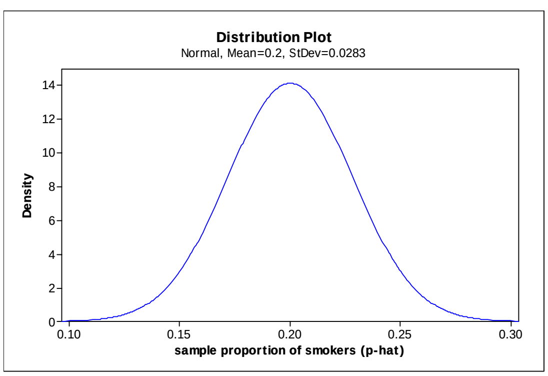
According to the Centers for Disease Control and Prevention, about 20% of all adult Americans are smokers. Let's assume that if you select an adult American at random, the probability is .2 that he/she is a smoker. Consider taking a random sample of 200 adult Americans. Denote the sample proportion of smokers by $\hat{p}$.

a) Verify that the Central Limit Theorem (CLT) applies here.

The population proportion of smokers is $\pi = .2$. The sample size is $n = 200$, so $n\pi = (200)(.2) = 40$ is larger than 10, and $n(1-\pi) = 200(.8) = 160$ is also larger than 10.

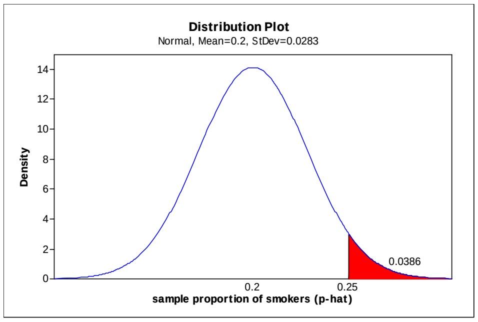
b) What does the CLT say about how the sample proportion of smokers would vary from sample to sample? [Be sure to comment on the shape, center, and spread of this sampling distribution. Also provide a well-labeled graph/sketch.]

The sample proportion of smokers, denoted by $\hat{p}$, would vary (approximately) according to a normal distribution with mean .2 and standard deviation $\sqrt{\frac{.2 \times .8}{200}} \approx .0283$, as shown here:



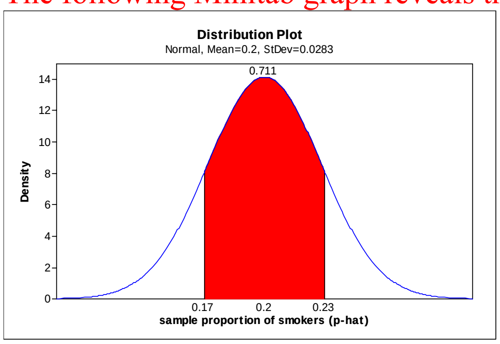
c) Use the CLT and the normal distribution to approximate the probability that more than 25% of the sample would be smokers.

The following Minitab graph reveals this probability to be approximately .0386:



d) Use the CLT and the normal distribution to approximate the probability that between 17% and 23% of the sample would be smokers.

The following Minitab graph reveals this probability to be approximately .711:

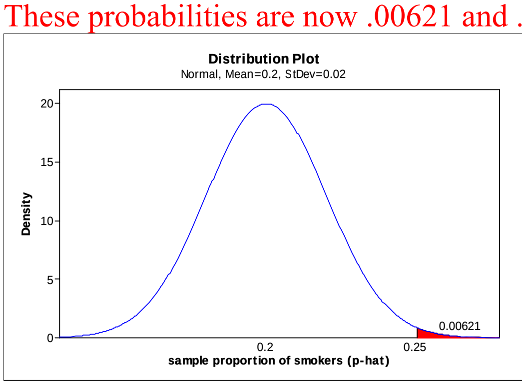


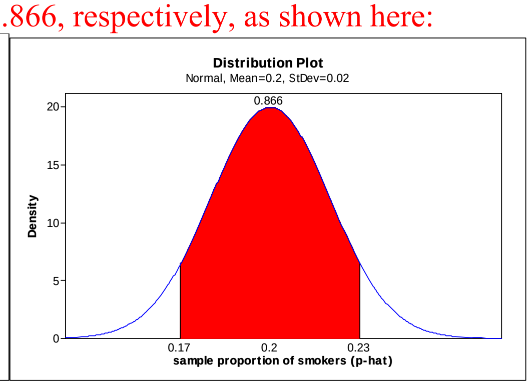
e) How would the sampling distribution of $\hat{p}$ (the sample proportion of smokers) change if the sample size were $n = 400$ instead of $n = 200$? What aspects of this sampling distribution remain unchanged?

The sampling distribution would still be approximately normal (in fact, with a larger sample size, it would be even closer to normal). The mean would still be .2. But the distribution would be less spread out, with a smaller standard deviation of $\sqrt{\frac{.2\times.8}{400}} \approx .0200$.

f) Recalculate your answers to (c) and (d) using a sample size of $n = 400$ rather than $n = 200$.

These probabilities are now .00621 and .866, respectively, as shown here:





g) Comment on how your answers to (c) and (d) changed with the larger sample size, and explain why these changes make intuitive sense.

The larger sample size reduces the variability in the sample proportions. So, the probability of obtaining a sample proportion in the tail (e.g., above .25) decreases, while the probability of obtaining a sample proportion near the mean (e.g., between .17 and .23) increases.