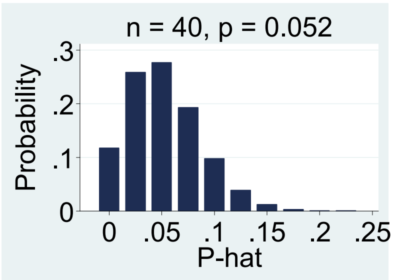
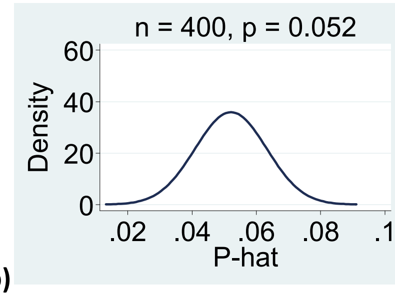
Homework 10: ECO220Y – SOLUTIONS

Required Problems:

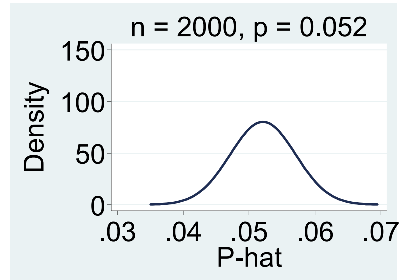
(1)



(a)



(b)



(c)

(d) For a sample size of 40, we should *not* use the Normal approximation. Hence, using the Binomial formula, we obtain $P(\hat{P} \leq 0.05 \mid n = 40, p = 0.0521) = P(X = 0) + P(X = 1) + P(X = 2) = 0.1176 + 0.2586 + 0.2772 = 0.6534$.

For sample sizes of 400 or 2,000, we can certainly use the Normal approximation using either of the two rules of thumb for the Normal approximation to the Binomial (e.g. we expect at least 10 successes and at least 10 failures).

$$P(\hat{P} \leq 0.05 \mid n = 400, p = 0.0521) = P\left(Z < \frac{0.05-0.0521}{\sqrt{\frac{0.0521*0.9479}{400}}}\right) = P(Z < -0.19) = 0.425$$

$$P(\hat{P} \leq 0.05 \mid n = 2{,}000, p = 0.0521) = P\left(Z < \frac{0.05-0.0521}{\sqrt{\frac{0.0521*0.9479}{2{,}000}}}\right) = P(Z < -0.42) = 0.34$$

For $n = 4{,}000$ we would have even less sampling error and the chance of getting a sample proportion as far below the true proportion ($p = 0.0521$) as 0.05 would become an even smaller probability than 0.34 above.

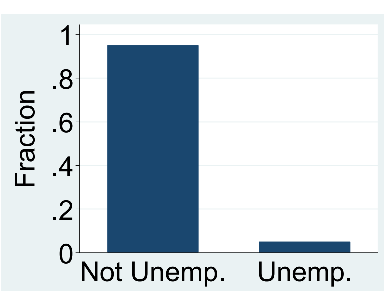
(e) Again, for a sample size of 40, we use the Binomial formula:
$P(\hat{P} \leq 0.025 \mid n = 40, p = 0.0521) = P(X = 0) + P(X = 1) = 0.1176 + 0.2586 = 0.3762.$

Again, for sample sizes of 400 or 2,000, we use the Normal approximation:

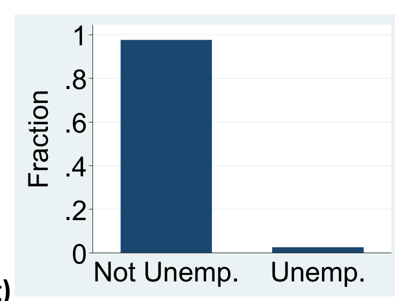
$$P(\hat{P} \leq 0.025 \mid n = 400, p = 0.0521) = P\left(Z < \frac{0.025-0.0521}{\sqrt{\frac{0.0521*0.9479}{400}}}\right) = P(Z < -2.44) = 0.007$$

$$P(\hat{P} \leq 0.025 \mid n = 2{,}000, p = 0.0521) = P\left(Z < \frac{0.025-0.0521}{\sqrt{\frac{0.0521*0.9479}{2{,}000}}}\right) = P(Z < -5.45) \approx 0$$

The probability for $n = 2{,}000$ is so tiny (basically zero) because this is a very large sample size and 0.025 is substantially below the true value of 0.0521: it is nearly impossible that a sample that big could yield a sample proportion that falls that far below the population proportion due to sampling error. Remember, as the sample size increases, sampling error goes down.



(f)



(g)

(2) (a) Following the four step method discussed in lecture:

Steps 1. and 2. yield the probability of every possible sample with three observations from the given population:

sample	probability	sample	probability	sample	probability
0,0,0	$(.4)^3$=.064	1,0,0	$(.4)^3$=.064	2,0,0	$(.2)(.4)^2$=.032
0,0,1	$(.4)^3$=.064	1,0,1	$(.4)^3$=.064	2,0,1	$(.2)(.4)^2$=.032
0,0,2	$(.2)(.4)^2$=.032	1,0,2	$(.2)(.4)^2$=.032	2,0,2	$(.2)^2(.4)$=.016
0,1,0	$(.4)^3$=.064	1,1,0	$(.4)^3$=.064	2,1,0	$(.2)(.4)^2$=.032
0,1,1	$(.4)^3$=.064	1,1,1	$(.4)^3$=.064	2,1,1	$(.2)(.4)^2$=.032
0,1,2	$(.2)(.4)^2$=.032	1,1,2	$(.2)(.4)^2$=.032	2,1,2	$(.2)^2(.4)$=.016
0,2,0	$(.2)(.4)^2$=.032	1,2,0	$(.2)(.4)^2$=.032	2,2,0	$(.2)^2(.4)$=.016
0,2,1	$(.2)(.4)^2$=.032	1,2,1	$(.2)(.4)^2$=.032	2,2,1	$(.2)^2(.4)$=.016
0,2,2	$(.2)^2(.4)$=.016	1,2,2	$(.2)^2(.4)$=.016	2,2,2	$(.2)^3$ =.008

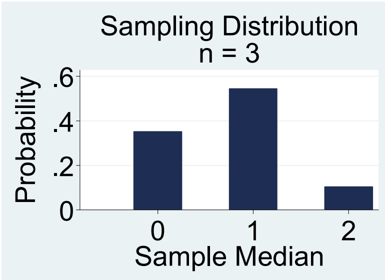
Step 3. yields the sample median of every possible sample:

sample	median	Sample	median	sample	median
0,0,0	0	1,0,0	0	2,0,0	0
0,0,1	0	1,0,1	1	2,0,1	1
0,0,2	0	1,0,2	1	2,0,2	2
0,1,0	0	1,1,0	1	2,1,0	1
0,1,1	1	1,1,1	1	2,1,1	1
0,1,2	1	1,1,2	1	2,1,2	2
0,2,0	0	1,2,0	1	2,2,0	2
0,2,1	1	1,2,1	1	2,2,1	2
0,2,2	2	1,2,2	2	2,2,2	2

Step 4. yields the sampling distribution of the sample median by combining the probability of every possible sample with the median of every possible sample:

Median	probability
0	0.352
1	0.544
2	0.104

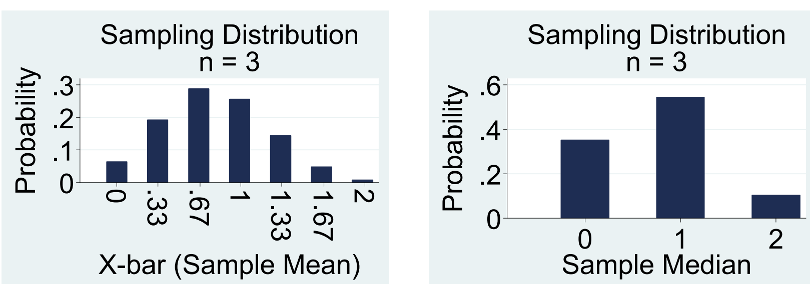
(b) The graph below show the sampling distribution of the sample median number of parking permits for a random sample of of 3 residents from the town. It shows that it is quite likely that the sample median will turn out to be 1: there is over a 50% chance of that. However, it is also fairly likely that the sample median may turn out to be 0: a bit over one-third of all samples would have a sample median on 0. It is also possible, but not as likely, that we may obtain a sample median of 2: there is about a 10 percent chance of that happening. Those are the ONLY three possibilities: given that every resident of the town has 0, 1 or 2 permits and given that we are randomly sampling three residents the median will be the middle value (2nd person) once we sorted the data by the number of permits: that middle person will have 0, 1, or 2 permits.



(c) $E[Median] = 0 * 0.352 + 1 * 0.544 + 2 * 0.104 = 0.752$
$V[Median] = (0 - 0.752)^2 0.352 + (1 - 0.752)^2 0.544 + (2 - 0.752)^2 0.104 = 0.394$
$SD[Median] = \sqrt{0.394} = 0.628$

To interpret these, if we *imagine* repeatedly drawing random samples of three residents from this town, on average the sample median will be 0.752. Notice that we are talking about an average of a sample statistic across samples. (Paraphrasing, a student in ECO220Y once said to me "Sampling distributions are hard because they involve statistics about statistics." That stuck with me as an interesting take on things.) Hence 0.752 makes sense even though the sample median of any given sample of three residents could never be 0.752 (it will be 0, 1, or 2). The reason that the sample median has a standard deviation is because of sampling error (i.e. sampling noise/sampling variability). Sometimes the sample median is 0, sometimes 1, and sometimes 2. The median of the population (a parameter) is 1: it does not have a (non-zero) standard deviation because parameters are constants and are not affected by sampling error (because a population is not a random sample). The standard deviation of the median – 0.628 – measures how much sampling error affects the sample median. This s.d. is big. That is not surprising, in a sample size of only 3 we would expect sampling error to be a big deal.

(d) The sampling distribution of the mean and median are quite different.



$E[\bar{X}] = 0.8$
$V[\bar{X}] = 0.187$
$SD[\bar{X}] = 0.432$

$E[Median] = 0.752$
$V[Median] = 0.394$
$SD[Median] = 0.628$

As a statistic, the sample median is more affected by sampling error than the sample mean in this parking permit example. That is clear both from the graphs and the fact that the standard deviation of the sample mean is smaller than the standard deviation of the sample median.

(e) Yes, right now things are entirely backwards. At this point in the course we are building up a theoretical foundation that will allow us to make inferences about populations using random samples a bit later in the course. But before we can get to that we are studying how sampling error would affect statistics given information about the population (which in reality we would not actually have).

(f) If the sample size were 30, step 1. would involve listing 3^{30} samples! If the town allowed each resident to have up to five permits (meaning that there are six possible values for a household: 0, 1, 2, 3, 4, 5), step 1. would involve listing 6^3 samples. Writing out the above tables with 27 possible samples (3^3) was already fairly tedious. In practice we will often rely on theoretical results that give us good approximations of the shapes of sampling distributions so long as some conditions are met. The reason we did this step-by-step method is to help you get a concrete grasp on the concept of a sampling distribution (often a stumbling block for students) before introducing more sophisticated theoretical results.

(3) (a) First, note that “as big as” means that big or bigger (a weak inequality). $P(\hat{P} \geq 0.60 \mid n = 200, p = 0.55) = P(X \geq 120) = P\left(Z > \frac{119.5-110}{\sqrt{200*0.55*45}}\right) = P(Z > 1.35) = 0.09$. *Note*: Recalling that $\hat{P} = \frac{X}{n}$, you could do everything in terms of $\hat{P}$. However, it is easier to think of X when doing the continuity correction.

(b) $P(\hat{P} \geq 0.60 \mid n = 200, p = 0.55) = P(X \geq 120) = P\left(Z > \frac{120-110}{\sqrt{200*0.55*45}}\right) = P(Z > 1.42) = 0.08$. *Note*: Recalling that $\hat{P} = \frac{X}{n}$, you could do everything in terms of $\hat{P}$ to obtain the exact same answer: $P(\hat{P} \geq 0.60) = P\left(Z > \frac{0.60-0.55}{\sqrt{\frac{0.55*0.45}{200}}}\right) = P(Z > 1.42) = 0.08$.