Sampling distribution of sample proportions

In these notes you will see how to apply the result known as the Central Limit Theorem to sample proportions. We first review the result. If a random sample is selected and the sample proportion \( \hat{p} \) of a particular outcome (which we call Success) is computed, then, provided the sample size \( n \) is large enough (usually above 30), and both \( np \geq 10, \ n(1- \ p) \geq 10 \), the distribution of sample proportions can be approximately modeled by a Normal model with mean \( p \) and standard deviation \( \sqrt{np(1-p)} \), where \( p \) is the true proportion of “successes” in the population.

Let us look at some examples.

**Exercise 5 chapter 18.** We are going to approach the problem using the Binomial model first. If the student obtained 42% heads in 200 tosses of a coin, it means the number of heads was actually 84. Assuming the tosses were independent, and the coin is fair, so that it has equal chance of landing heads or tails, the number of heads in 200 tosses follows a Binomial model with \( n = 200, p = 0.5 \). Then in 200 tosses we expect to obtain about 100 heads. Now we know that due to variability the actual number may not be exactly 100. The question we need to answer is whether 84 heads is unusually small so that what we are witnessing is an unlikely event. In order to answer the question we will compute the probability of obtaining 84 or even fewer heads in 200 tosses. So if \( X \) is the random variable counting the number of heads we compute:

\[
P(X \leq 84) = \text{binomcdf}(200, 0.5, 84) = 0.0141.\]

This result can be interpreted like this: 84 heads or fewer can be observed in about 1.4% of such experiments. Since the probability of the event is so small (much below 0.05) we may conclude either of the following:

(a) We are witnessing a very unusual event;
(b) The coin may not be fair, so in reality the true proportion of heads may be below 0.5,
(c) That student may have made a mistake in counting the number of heads.

Now we approach the same problem using the sampling distribution of the sample proportion of the number of heads. Assuming again that the coin is fair, \( p = 0.5 \) is the true proportion of heads. In 200 tosses we expect about 100 heads and 100 tails, so the conditions are satisfied and we can use the normal model for the sample proportion. Again, we would like to assess whether \( \hat{p} = .42 \) is unusually small.

According to the Central Limit Theorem, the sample proportions follow a Normal model with mean \( p = 0.5 \) and standard deviation \( \sigma = \sqrt{\frac{0.5 \cdot 0.5}{200}} = 0.035 \). Just for a rough estimate, we can see that .42 is more than two standard deviations below the mean, so this is enough to tell us this is an unusual event to observe. We will be more precise and compute the probability of observing such a low proportion of heads or even lower
than that, so we will compute \( P(X \leq 0.42) \) for a normal model with mean \( p = 0.5 \) and standard deviation \( \sigma = \sqrt{\frac{0.5 \cdot 0.5}{200}} = 0.035 \).

We have \( P(\hat{p} \leq 0.42) = \text{normalcdf}(0, 0.42, 0.5, 0.035) = 0.0111 \).

Notice that even if the probability we computed is not the same as the one we obtained using the Binomial model, the two values lead to the same conclusion: the probability is very small, so this is a very unlikely event.

**Note:** Review the TI83 command for computing probabilities in the Normal model. Remember that in Chapter 6 when we computed normal probabilities, the probability of the event above was computed as:

\[
P(\hat{p} \leq 0.42) = \text{normalcdf}(-E99, 0.42, 0.5, 0.035) = 0.0111
\]

Notice that the value is identical to the one computed above, and the reason is that 0 is sufficiently far (more than 3 standard deviations) from the mean 0.5 so that any area to the left of zero is practically zero. Also, it makes more sense from the point of view of computations to look at values between 0 and 0.42 rather than \(-\infty\) and 0.42 since the sample proportion can only take values between 0 and 1.

**Exercise 14.** Let us compute the sample proportion of students engaged in binge drinking at that particular college. We have \( \hat{p} = \frac{96}{244} = 0.393 \). Compared to 44% it falls below that value. The question we would like to answer is whereas this proportion is much smaller, or unusually smaller compared to 44%.

We first check that the conditions of using the normal model apply. We have a large sample, \( n=244 \), and both \( np = 244 \cdot 0.44 = 107.36 \geq 10 \), and \( n(1-p) = 136.64 \geq 10 \).

So \( \hat{p} \) follows a Normal model with mean 0.44 and standard deviation \( \sigma = \sqrt{\frac{0.44 \cdot 0.56}{244}} = 0.031 \). We have: \( P(\hat{p} \leq 0.393) = \text{normalcdf}(0, 0.393, 0.44, 0.031) = 0.0647 \).

So, the probability of having a sample in which the proportion of binge drinking students is 0.393 or even lower is 0.0647. This is not a large probability, but according to our rule for unusual events, the probability is above 0.05, so we cannot classify the event as an unlikely one.

Therefore, we should not be very surprised at this result.

Again, let us approach the same problem using the Binomial model. We can assume that the number of students who binge drink follows a Binomial model with \( n=244 \) and probability of “success” \( p=0.44 \). Then, to answer the question in the problem we compute \( P(X \leq 96) = \text{binomcdf}(244, 0.44, 96) = 0.0802 \). The probability is not equal to the one obtained using the Normal model, however, since the value is 0.08>0.05, again we conclude the event is not unusual, so we should not be surprised.

**Exercise 20.** First of all, in this problem the answers may vary quite a lot based on everybody’s understanding of the word “very sure”. So, if we look at the proportion
\( \hat{p} \) of people who will order the chef’s steak special, it follows a model which is approximately Normal with mean 0.20 and standard deviation

\[
\sigma = \sqrt{\frac{0.2 \cdot 0.8}{180}} = 0.0298 \sim 0.03
\]

Arguing that values that are farther than two standard deviation above the mean are unusual, it means it would be unusual to have more than 0.20+0.06=0.26=26% orders for the chef’s special.

So, if I were to make a recommendation, I would tell the chef to prepare 180 \cdot 0.26 \approx 47 \) chef’s special steaks.

Now, to play it even safer, we may prepare the chef’s special for 27.5% (i.e. two and a half standard deviations above the mean) of the expected number of customers, this leading to preparing 50 chef’s special steaks.

Now, if you open this document and read it, work problems 18 and 19 following the examples I provided. You need not give two solutions as I gave for exercises 5 and 14, but I would like to see complete solutions with explanations.

Write the solutions nicely on paper and bring them on Monday to class for extra credit. You are allowed to work together and discuss the problems but make sure to show your individual work and explanations.