Sampling Distributions and Inference
What Do We Know and How Do We Know It?
(Teacher Notes)

This handout describes a set of simulation activities suggested for use after students have studied probability, including binomial distributions, and before being introduced to formal inference procedures. All graphs were generated using TI-Interactive! software.

The primary goals of these simulations are to make clear that sampling distributions summarize the set of possible sampling results for a given sampling process, and to help develop the notion that for each value of a population parameter there is an associated characteristic sampling distribution for a statistic. As the population changes, so does the sampling distribution.

Additional goals are to introduce the concepts of hypothesis tests and confidence intervals and to relate these two types of inference to each other. A hypothesis test asks whether a particular observed sample statistic might reasonably have come from a population having a hypothesized parameter value. A confidence interval lists all "reasonable" values of the population parameter. Thus it may be thought of as the set of all parameter values for which the corresponding hypothesis tests fail to reject H_0.

The following sequence of activities is restricted to populations consisting of only two types of elements, "success" and "failure". The goal of the inference is to estimate the proportion, p, of the population that are successes.

Problem Setting:

Suppose that an automobile manufacturer wishes to have 60% of its vehicles be free of warranty-related repairs during the first three years after the car is sold. If too few cars go three years without repairs, the company will be branded as selling "lemons." If too many cars need no repairs, opportunities for income through the service department of dealerships diminishes. One dealership decides to check this performance by examining the service records of 20 randomly selected cars sold three years ago. They find that exactly 5 of these 20 cars have needed no repairs covered under warranty. Is this evidence that the proportion of <u>all</u> cars needing warranty repairs during the first three years is something other than 60%? What is a reasonable estimate of the actual percentage of all cars sold that need no covered repairs during the three-year period?

The dealer needs a statistical estimate of the proportion, p, of cars reaching the three-year term without requiring repairs under warranty. We will examine this scenario through a sequence of simulations, building from physical to numerical to calculator or computer based. The primary goal is to see what "should" happen for known populations (values of p). The resulting sampling distributions will then permit estimation of the actual (but unknown) proportion p.

Physical Simulation:

For the physical simulation, provide each student a bag containing a mix of candies, some red and some not; red candies will represent "success." Each bag represents a population, and the proportion of candies in that bag that are red is the population parameter, p. Here, of course, we *can* determine p; all we need to do is count. Label the bottom of each bag with its respective p. (Use a wide range of values of p, from as low as 0.10 to at least 0.90. Be sure to include a bag with $p = 0.60$.)

Have students select (with replacement) samples of size 20 from their bags, recording the number, x, of candies in each sample that are red. Repeat the sampling so that each student obtains several samples, thus several x values, from which a histogram (or dot plot, or . . .) can be formed.

Initially, students should decide whether the outcome of $x = 5$ is likely or unlikely to occur in samples of size 20 from their assigned populations. Later they should begin to partition the set of *all* values of x into likely and unlikely values for their population. Discuss these ideas informally after each simulation, using $p = 0.6$ (for the car scenario) and a few others for comparison and generalization. Note, too, that although $x = 5$ is on the "low side" of $p = 0.6$, it would also be a problem for the dealer if x appeared too high. Unlikely values can occur at either end of a distribution.

Physical simulation has a couple of major drawbacks. First, it's slow. It takes time to complete enough runs for the distribution of x to be informative. Ten runs probably are not enough to get a sense of what all possibilities might be.

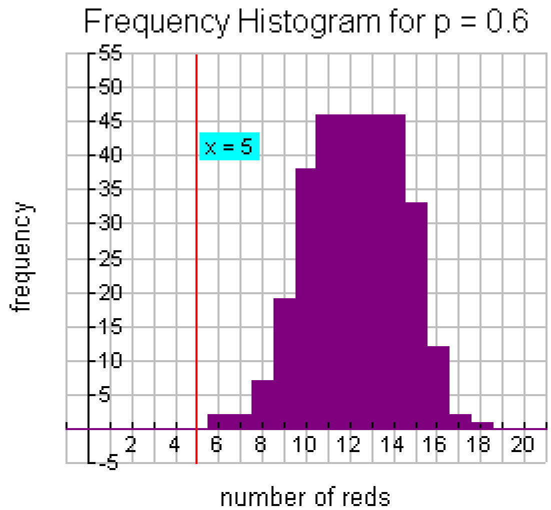
Even worse, though, is the lack of flexibility in selecting values of p to use in the simulation. Each new value of p that you wish to examine requires a new bag, carefully counted to contain the correct proportion. This limits the scope of the inference. Numerical simulation addresses these problems.

Numerical and Calculator Simulations:

Use a calculator to select 20 random uniform values from [0, 1), assigning [0, p] as "success" and (p, 1) as "failure". (Other procedures will do equally well.) Count the number of successes in the sample of 20.

(For TI-83 calculators, **rand(20)** will produce the desired sample, which may be stored to a list, say L1. Then the expression **sum(L1 ≤ 0.6)** returns a count of the number of entries in L1 that are less than or equal to 0.6; that is, it gives x if $p = 0.6$. The command "sum" adds values of a list. Its argument, "L1 ≤ 0.6" tests each element of L1 and reports the value "1" if the condition is true.)

Once students are comfortable using calculators to implement the numerical simulation, they are ready to move to more efficient mechanisms, taking full advantage of available technology. For example, some computer software and calculators include commands to generate random binomial values, from which a histogram or other graph may be made. Here are example output based on 300 random x values from a B(20,0.6) population. The observed $x = 5$ is marked in the histogram.

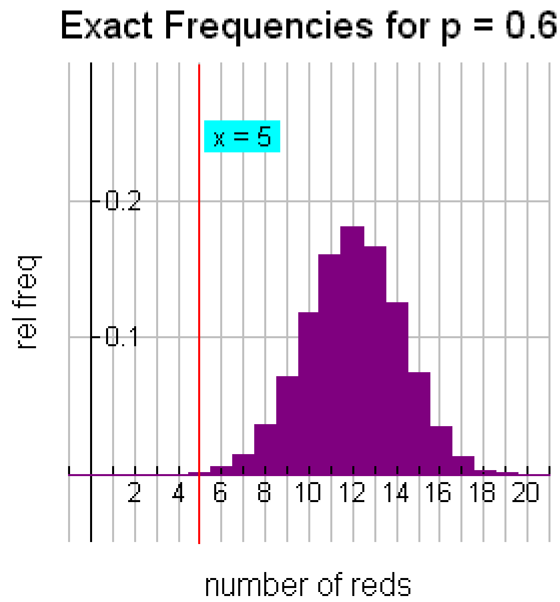


For these simulated data, having 5 or fewer good cars (successes) failed to occur within the 300 runs. Thus it appears that seeing 5 or fewer successes in a sample of 20 from a population with $p = 0.6$ is a very rare event. We would estimate its probability as being less than 1% (i.e., in less than 1% of all possible samples).

The Logic of Inference:

Each student has a different population (value of p). Thus, each student has a different simulated histogram of x values. The previous simulation asked students to use their histograms to estimate the probability of the event "$x \leq 5$" for a random sample of size 20 taken from their own populations. To eliminate errors due to the approximate nature of simulations, simulated results may be replaced by exact sampling distributions using formulas for the binomial probability distribution function. An example of exact data is shown below.

success	proba
0	1E-008
1	3E-007
2	5E-006
3	4E-005
4	0.00027
5	0.00129
6	0.00485
7	0.01456
8	0.0355
9	0.07099
10	0.11714
11	0.15974
12	0.17971
13	0.16588
14	0.12441



It is highly instructive to have students display their respective sampling distributions, together with the vertical marker at the observed sample (x = 5), side by side along a long table or chalkboard tray, in number-line order (for example, 0.05, 0.10, 0.15, . . ., 0.95). Label each graph with the value of the corresponding population p. For chalkboards, it is easy to list these values at about 6-inch intervals about a foot above the chalk tray. Before arranging the graphs, determine appropriate scales so that all graphs may be displayed using identical settings for easier and more informative comparisons.

Following the comparison, have students retrieve their graphs. Help the class to agree on a common criterion by which to conclude "I don't believe that observation came from my population". (Typically, students are comfortable rejecting the outermost 10% of a distribution as being relatively unlikely. That is, observations in the left- or right-most 5% of the distribution are deemed unlikely.) Using that value, each student may reject or fail to reject the claim that the observed sample came from their own particular bag, depending on whether the vertical line at the observed value falls within the rejection region or not. This corresponds to the traditional two-sided hypothesis test, with $H_0 : p$ = (assigned value) at the 0.10 significance level.

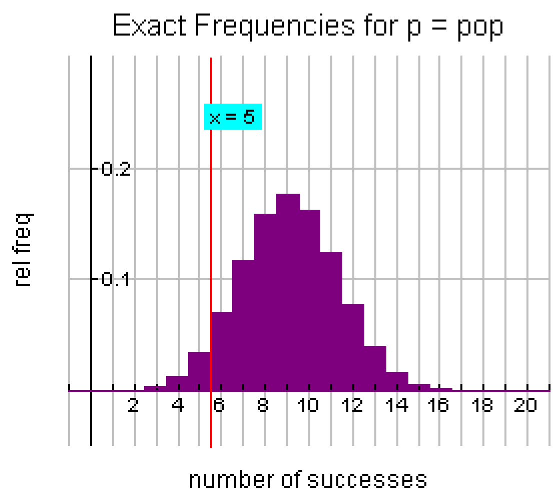
Using the agreed-upon rejection criterion, have all students carry out the hypothesis test using the observed x = 5. The set of students who fail to reject H_0 for the observed sample determine an approximate (90%, using the criteria described above) confidence interval for p from the observed x. Thus, a candidate population p is in this interval only in case the observed x is "likely" to have come from that population. (The approximation arises due to the inability to test all values of p within [0, 1].)

TI-Interactive!

After the "low-tech" cartoon of lined up calculators, the set of graphs of sampling distributions associated with the set of all populations may be viewed dynamically in TI-Interactive! A "slider" or a "math box" gives the user the ability to select the desired *p* directly or to change values smoothly.

$pop := .455$

success	proba
0	5E-006
1	9E-005
2	0.00071
3	0.00355
4	0.01258
5	0.03361
6	0.07015
7	0.11712
8	0.1589
9	0.17688
10	0.16243
11	0.12328
12	0.07719



For a given null *p*, the exact p-value for a two-sided alternative when the observed $x < np$, using binomial probabilities is:

$$2 \cdot \text{binomcdf}(20, pop, 5) = .101073$$

For observed $x > np$, the corresponding p-value is given by:

$$2 \cdot (1 - \text{binomcdf}(20, pop, 4)) = 1.96614$$

Of course, only one of the two p-value equations will make sense in any given situation!

Note that $p = 0.455$ is the highest (to three figures) parameter value for which $x = 5$ falls within the central 90% of the sampling distribution. It is therefore the right endpoint of the 90% confidence interval for *p* based on a sample with $x = 5$. The corresponding left endpoint is $p = 0.104$.