Activity: Hypergeometric Probabilities

Concepts: Applying Counting Techniques, Hypergeometric Probabilities

Prerequisites: Knowledge of basic counting techniques (permutations and combinations), computing probabilities using the equally likely principle.

Review of Counting Techniques:

Recall the *general product rule*: If a process consists of multiple stages (call the number of stages k) and if stage i can be completed in n_i ways regardless of which outcomes occur in earlier stages, then the process itself can be completed in $n_1 n_2 \cdots n_k$ ways.

Recall the number of *permutations* (where order matters) of n objects taken k at a time is: $n(n-1)(n-2)\cdots(n-k+1)$. Notice that there are k terms in this product. This can also be expressed as: $P_{k,n} = \frac{n!}{(n-k)!}$.

Recall the number of *combinations* (where order does not matter) of n objects taken k at a time is: $C_{k,n} = \binom{n}{k} = \frac{n!}{(n-k)!k!}$. This symbol $\binom{n}{k}$ is often pronounced "n choose k" because it is the number of ways of selecting a group of k objects from a collection of n objects.

Examples (Permutation vs. Combination):

1. A computer system has 20 circuits. How many ways are there to randomly select 6 circuits?
2. Three married couples purchase theater tickets, 6 seats in a row. How many ways are there for them to arrange themselves? How many ways can they be arranged if each individual couple does not want to be separated?
3. How many different five-card stud poker hands are there?
4. A radio station manager can choose between R&B, rock and roll, and classical music. How many different arrangements are there?
5. How many ways are there to have 3 R&B songs in the next 5 songs?

Scenario: College Committee Formation I

A college professor found herself to be one of two women on a six-person committee. The committee needed to choose two representatives to speak for the committee to other groups, and both women were selected. You will investigate how unusual this would be if the two representatives had in fact been chosen at random.

(a) Let the names of the committee members be: Alice, Barb, Chuck, Dave, Ethan, and Fred. Using initials to represent these people, list the sample space (set of all possible pairs) for this process.

(b) How many outcomes are in this sample space? What counting rule above could you have used to find this number? Explain and verify.

(c) If the two representatives had been chosen at random (without regard to gender), all possible pairs of the six members would have been equally likely. What then is the probability for each of the possible pairs?

(d) Assuming equal likeliness, what is the probability that both persons selected would be women? What counting technique could you have used to find the numerator in this fraction? Explain and verify.

(e) Does this probability indicate that it would be very surprising to find both representatives women if all possible pairs were equally likely? Explain.

(f) Assuming equal likeliness, what is the probability that both persons selected would be men? See if you can find this probability by using counting techniques.

(g) Assuming equal likeliness, what is the probability that one person of each gender would be selected? Again see if you can also find this probability by using counting techniques.

(h) What do you notice about the sum of the probabilities in (d), (f), and (g)? Explain why this makes sense.

(i) Now suppose that the committee had been choosing a chair and a secretary rather than just two representatives. In other words, suppose now that order matters, so Alice as chair and Barb as secretary is a different outcome than Barb as chair and Alice as secretary. Without listing them, indicate how many outcomes comprise this sample space. How does this number compare to the previous one where they were simply two representatives? What counting technique above could you have used to find this number? Explain why this makes sense.

(j) In this new scenario, how many of the outcomes consist of two women? What is the probability that the chair and secretary would both be women? Has this probability changed from the one calculated with the original sample space (where order did not matter)?

In many probability calculations, it does not matter whether you count as if order matters or not, as long as you *count consistently* in both the numerator and denominator of the probability calculation. In this case, the probability of choosing two women is 1/15 or 2/30 depending on how you count.

Hypergeometric Probabilities:

Suppose that a *population* of N objects consists of r "successes" and N-r "failures." Suppose that a *sample* of n objects is to be chosen at random. Consider the probability of drawing x successes in the sample.

(k) How many ways are there to choose n objects from the population of N objects? Write your answer in terms of n and N.

(l) How many ways are there to select x successes from the population of r successes? Write your answer in terms of x and r.

(m) How many ways are there to select n-x failures from the population of N-r failures? Write your answer in terms of n-x and N-r.

(n) Produce an expression for the probability of obtaining x successes in the sample of n objects (Hint: Multiply your answers to (l) and (m) and then dividing by your answer to (k)).

You should have found that the probability of drawing x successes in a sample of size n from this population is: $\dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$. These are called *hypergeometric* probabilities.

Return to considering the population of four men and two women, with a sample of size two to be drawn from that population.

(o) Use the above expression for hypergeometric probabilities to find the probability of selecting two women. Then do the same for the probability of choosing one woman. Then repeat for zero women. Do these agree with your calculations in (d), (f), and (g)?

(p) Use Minitab or your calculator to verify these calculations. To use Minitab, put the values 0, 1, and 2 into `c1` and then select `Calc > Probability distributions > Hypergeometric`. Be sure to click on the "Probability" option and enter the appropriate parameter values.

<u>Scenario:</u> Friendly Observers II

Reconsider the psychology experiment about whether an observer with a vested interest inhibits one's performance on a skilled task. To assess whether the experimental results (3/12 successes in group A vs. 8/12 successes in group B) were unlikely to occur just by random chance, you performed a simulation analysis. Now you are ready to calculate this probability exactly using counting techniques, specifically hypergeometric probabilities.

(q) Specify the relevant hypergeometric parameters N, r, and n for calculating the probability of finding three or fewer successes in group A just by chance.

(r) Calculate the probability in (q) either by hand or by using Minitab, i.e. find the p-value of the experimental results. Comment on how closely your empirical estimates of this probability came to the actual value. Finally, summarize whether this probability suggests that the experimental results are very unlikely to have occurred by chance alone if there were no effect of the observer with the vested interest. Use a probability cutoff of 0.05 or less as being very unlikely to have occurred by chance alone.