### Independence Requirement

When selecting a sample (such as survey subjects) for some statistical analysis, we usually sample without replacement. Sampling without replacement involves dependent events, which violates the second requirement in the definition above. However, we can often assume independence by applying the following 5% guideline introduced in Section 4-4:

**Treating Dependent Events as Independent: The 5% Guideline for Cumbersome Calculations**

**If calculations are cumbersome and if a sample size is no more than 5% of the size of the population, treat the selections as being *independent* (even if the selections are made without replacement, so that they are actually dependent).**

If a procedure satisfies the four requirements above, the distribution of the random variable $x$ (number of successes in $n$ trials) is called a *binomial probability distribution* (or *binomial distribution*). The following notation is commonly used.

**Notation for Binomial Probability Distributions**

S and F (success and failure) denote the two possible categories of all outcomes.

| | |
|---|---|
| $P(\text{S}) = p$ | ($p$ = probability of a success) |
| $P(\text{F}) = 1 - p = q$ | ($q$ = probability of a failure) |
| $n$ | denotes the fixed number of trials. |
| $x$ | denotes a specific number of successes in $n$ trials, so $x$ can be any whole number between 0 and $n$, inclusive. |
| $p$ | denotes the probability of *success* in *one* of the $n$ trials. |
| $q$ | denotes the probability of *failure* in *one* of the $n$ trials. |
| $P(x)$ | denotes the probability of getting exactly $x$ successes among the $n$ trials. |

The word *success* as used here is arbitrary and does not necessarily represent something good. Either of the two possible categories may be called the success S as long as its probability is identified as $p$. (The value of $q$ can always be found by subtracting $p$ from 1: If $p = 0.95$, then $q = 1 - 0.95 = 0.05$.)

**CAUTION** When using a binomial probability distribution, always be sure that $x$ and $p$ are consistent in the sense that they both refer to the *same* category being called a success.

### Example 1 Twitter

When an adult is randomly selected (with replacement), there is a 0.85 probability that this person knows what Twitter is (based on results from a Pew Research Center survey). Suppose that we want to find the probability that exactly three of five random adults know what Twitter is.

**a.** Does this procedure result in a binomial distribution?

**b.** If this procedure does result in a binomial distribution, identify the values of $n$, $x$, $p$, and $q$.

### Not at Home

Pollsters cannot simply ignore those who were not at home when they were called the first time. One solution is to make repeated callback attempts until the person can be reached. Alfred Politz and Willard Simmons describe a way to compensate for those missing results without making repeated callbacks. They suggest weighting results based on how often people are not at home. For example, a person at home only two days out of six will have a 2/6 or 1/3 probability of being at home when called the first time. When such a person is reached the first time, his or her results are weighted to count three times as much as someone who is always home. This weighting is a compensation for the other similar people who are home two days out of six and were not at home when called the first time. This clever solution was first presented in 1949.