

Sampling Distributions

Sampling Distributions

A statistic from a random sample or randomized experiment is a *random variable*. The probability distribution of the statistic is its **sampling distribution**.

Population Distributions

The **population distribution** of a variable is the distribution of its values for all members of the population. The population distribution is also the probability distribution of the variable when we choose one individual from the population at random.

Sampling Distribution for Counts and Proportions

Count (X) - number of occurrences of some outcome in a *fixed* number of observations.

Sample Proportion - number of occurrences out of the number of observations.

$$\hat{p} = X/n$$

The Binomial Distribution

The Binomial Setting

1. There are a fixed number n of observations.
2. The n observations are all independent.
3. Each of the observations falls into one of just two categories, which for convenience we call “success” and “failure.”
4. The probability of a success, call it p , is the same for each observation.

Binomial Distribution: The distribution of the count X of successes in the binomial setting is called the binomial distribution with parameters n and p .

- n is the number of observations
- p is the probability of a success on any one observation
- The possible values of X are whole numbers from 0 to n .

Notation: X is $B(n,p)$

For each of the following situations, indicate whether the binomial distribution is a reasonable probability model for the random variable X .

- a) You observe the sex of the next 50 children born at a local hospital; X is the number of girls among them.

- b) A couple decides to continue to have children until their first girl is born; X is the total number of children the couple has.

-
- c) A chemist repeats a solubility test 10 times on the same substance. Each test is conducted at a temperature 10 degrees higher than the previous test. She counts the number of times that the substance dissolves.

 - d) Fifty students are taught about the binomial distribution in class. They are then given a short quiz on the subject. The number of students who pass is counted.

The Binomial Distribution as a Sampling Distribution

We use the binomial distribution in statistical inference when we have count data (in the binomial setting) and we want to make inferences about the total proportion of “successes” in a population.

When the population is much larger than the sample, the count X of successes in a SRS of size n has *approximately* a $B(n,p)$ distribution if the population proportion of successes is p .

Rule of thumb: use the binomial sampling distribution for counts when the population is at least 10 times larger than the sample.