

ST 518: Data Analytics II

Binomial Random Variables

Binomial Random Variables

Binomial Counts

The Binomial Distribution

Example

Binomial Counts

Suppose we have a binary outcome:

$$Z = \begin{cases} 1 & \text{if event } E \text{ occurs} \\ 0 & \text{if event } E \text{ does not occur} \end{cases}$$

and that E occurs with probability, p . Using statistical notation we write:

$$Z \sim \text{Bernoulli}(p).$$

Further suppose that we have the resources to run n experiments or trials for which we can observe whether or not E occurs.

A binomial random variable is a count of the number of times E occurs in the n trials.

Examples

- Medicine: The number of patients in a clinical trial that showed improvement on a new treatment.
- Politics: The number of voters in an exit poll sample who said they voted for Hillary Clinton.
- Ecology: The number of wetlands that have seen an increase in the number of bird species.
- Business: The number of \$5 off coupons redeemed on a customer appreciation day.
- Health: The number of women with heart disease among a sample of obese women.

The Binomial Distribution

Mathematically, a binomial random variable is defined as the sum of n independent Bernoulli random variables, all with probability p . Suppose

$$Z_1, Z_2, \dots, Z_n \stackrel{ind}{\sim} \text{Bernoulli}(p)$$

then

$$X = Z_1 + Z_2 + \dots + Z_n = \sum_{i=1}^n Z_i \sim \text{bin}(n, p)$$

The binomial probability distribution, or probability mass function is

$$\Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

where $x \in \{0, 1, \dots, n\}$, $p \in [0, 1]$ and the term in front, $\binom{n}{x}$, is called the binomial coefficient.

The Binomial Distribution

The binomial coefficient is a combinatorial term that counts up the number of ways x things can be chosen out of n things.

- The mathematical details are not that important for our purposes
- You can use R to calculate binomial coefficients (see the function **choose**)
- And, you can use R to calculate properties of binomial random variables—for instance their probability and/or their cumulative probability (see the functions **dbinom** and **pnbinom**)

Vitamin C Example

In a study reported by Ramsey & Schafer (2013), 818 volunteers were randomly assigned to take 1000mg of Vitamin C or a placebo throughout the cold season. At the end of the cold season, each subject was evaluated for whether he or she had a cold. Here are the data:

	<u>Outcome</u>		Totals
	Cold	No Cold	
Placebo	335	76	411
Vitamin C	302	105	407
Totals	637	181	818

You can access these data in R by installing and loading the “Sleuth3” library and looking at the **case1802** object.

Vitamin C Example

For now, let's just focus on the 411 subjects who were assigned to take the placebo. At the end of the cold season, 335 had colds.

1. What is n in this situation?

Vitamin C Example

For now, let's just focus on the 411 subjects who were assigned to take the placebo. At the end of the cold season, 335 had colds.

1. What is n in this situation?
2. What is the value of the binomial count, X ?

Vitamin C Example

For now, let's just focus on the 411 subjects who were assigned to take the placebo. At the end of the cold season, 335 had colds.

1. What is n in this situation?
2. What is the value of the binomial count, X ?
3. What's our best guess at p , the proportion of subjects on placebo who get colds?