Unit 5: Introduction to probability

2. Random variables and Binomial distribution

Stat 140 - 02

Mount Holyoke College

Dr. Shan Shan

Slides posted at *http://sshanshans.github.io/stat140*

1. Motivation: friend or foe

2. Today: random variable and binomial distribution

3. Main ideas
   1. Definition of a random variable
   2. Permutations and Combinations
   3. Binomial Distribution

4. Summary

10-month-old infants were shown a “climber” character (a piece of wood with “googly” eyes glued onto it) that could not make it up a hill in two tries. Then they were alternately shown two scenarios for the climber’s next try, one where the climber was pushed to the top of the hill by another character (“friend”) and one where the climber was pushed back down the hill by another character (“foe”).

The researchers found that 14 of the 16 infants in the study selected the nice toy.

Do the infants have preference to “friend” toy over “foe”?

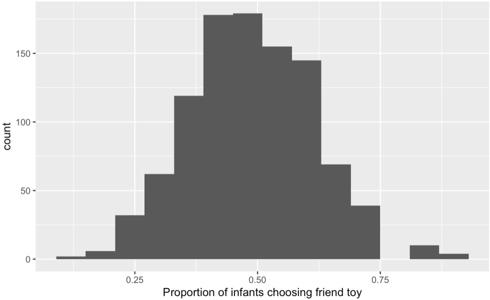
Identify population parameter and sample statistic

- Population parameter:
  $p =$ the proportion of infants select the "friend" toy among all infants
- Sample statistic:
  $\hat{p} =$ the proportion of infants select the "friend" toy among the 16 infants in the study

State hypothesis

- Null hypothesis (H_0): $p = 0.5$
- Alternative hypothesis (H_A): $p > 0.5$

Simple sampling distribution of sample statistic, assuming H_0 is true



Calculate the p-value for the test

Find those samples with a proportion of infants selecting the friend toy that is greater than $14/16 = 0.877$, and that is 0.001. Since the p-value is less than 0.05, we reject the null hypothesis.

The key is to compute the p-value

$P(\text{data} \mid H_0) = P(\text{14 out of 16 infants select friend toy} \mid p = 0.5)$

Some abstraction ...

- The frequentist notion of probability: $p = 0.5$
  proportion $\rightarrow$ probability an infant select the friend toy
- Let Y denote the number of infants in a sample of size 16 who choose a helpful toy.
- We observed a sample with 14 infants select a helpful toy

We need to compute

$$P(Y = 14 \mid p = 0.5)$$

outcome:

- choose friend toy
- choose foe toy

$p = 0.5$:
probability of an infant select the friend toy

Y:
# of infants in a sample of size 16 who choose a helpful toy.

compute:

$$P(Y = 14 \mid p = 0.5)$$

outcome:

- head
- tail

$p = 0.5$:
probability of a coin toss results in a head

Y:
# of heads in 16 tosses of a coin

compute:

$$P(Y = 14 \mid p = 0.5)$$

Probability of exactly 14 heads in 16 tosses of a coin

A **random variable** is a variable whose possible values are numerical outcomes of a random phenomenon. Use capital letters like X, Y, Z to denote random variables. Use lower case letters x, y, z to denote specific observed values.

Here are some examples:

- Let X denote the number of heads in 10 tosses of a fair coin for. Maybe in a particular sample I observe $x = 3$.
- Let Y denote the number of infants in a sample of size 16 who choose a helpful toy. In our sample we observed $y = 14$.
- Let Z denote the number of M&M's in a sample of size 100 that are blue. Maybe in a particular sample I observe $z = 22$.

Remember our goal is to figure out $P(Y = y)$.

The number of ways to arrange n distinct objects in a line is

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 1.$$

Here $n!$ is called **n factorial** and it is the number of permutations of n distinct objects.

There are n choices for the first position, then $n-1$ for the second, and so forth; multiplication gives the number of distinct arrangements.

By convention, $0! = 1$. For the other positive integers,

$$1! = 1, \quad 2! = 2, \quad 3! = 6, \quad 4! = 24, \ldots$$

In how many ways can 8 people line up? $8! = 40320$.

In how many ways can 4 married couples stand in a police line-up, if couples must stand together? There are $4! = 24$ ways that the couples can be arranged, and each couple can be arranged in $2! = 2$ ways. So the answer is

$$(4!)(2!)(2!)(2!)(2!) = 384.$$

The number $\binom{n}{r}$ is also the number of ways to pick r objects from a set of n distinct objects.

$$\binom{n}{r} = \frac{n!}{r! \times (n-r)!}$$

For example, when I was a kid, I had 20 stuffed animals. My bed could only hold 7. Being perfectly promiscuous, I chose a different set each night. How long before I must repeat a set?

$$\binom{20}{7} = \frac{20!}{7! \times 13!} = 77520$$

That's about 212 years.

The binomial formula gives the probability of exactly x successes in n tries, where each try has the same probability of success p and each try is independent.

$$P[\text{ exactly x successes }] = p(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

This is exactly the situation one has when trying to find the probability of x heads in n tosses of a coin that has probability p of coming up heads.

This random variable X has the binomial distribution $\text{Bin}(n, p)$.

Why does this formula work?

1. How many arrangements are there that give x heads in n tries? From the previous section, we know the answer is $\binom{n}{x}$
2. Each arrangement is incompatible (disjoint) with the other arrangements. Thus the probability of exactly x successes is the sum over all possible arrangements, and there are $\binom{n}{x}$ of those.
3. Each arrangement has the same probability: $p^x(1-p)^{n-x}$. To see this, consider the sequences $HTHT$ and $THTH$. The first has probability $p(1-p)p(1-p) = p^2(1-p)^2$. The second arrangement has probability $(1-p)p(1-p)p = p^2(1-p)^2$.

1. $X \sim \text{Bin}(n, p)$: X follows a Binomial distribution with sample size n and probability of success p.
2. $P(X = x)$: the probability that $X = x$, for each possible value of x from 0 to n.
3. Binomial formula:

$$P[X = x] = \binom{n}{x} p^x (1 - x)^{x-n}$$

Find the probability of exactly two sixes in five rolls of a fair die.

1. Motivation: friend or foe
2. Today: random variable and binomial distribution
3. Main ideas
   1. Definition of a random variable
   2. Permutations and Combinations
   3. Binomial Distribution
4. Summary

1. Definition of a random variable
2. Permutations and Combinations