

22S:105
Statistical Methods and
Computing

Introduction to Probability

Lecture 9
February 25, 2008

Kate Cowles
374 SH, 335-0727
kcowles@stat.uiowa.edu

Parameters and statistics

- A **parameter** is a numeric quantity that describes a characteristic of a population.
 - We almost never can know the exact value of a parameter, because we would have to measure every member of the population.
 - Example: We would like to know the average percent body fat of all Chinese males aged 21 - 65 years.
 - We generally use Greek letters to refer to population parameters.
 - μ is the standard symbol for a population mean.

Why do we want to study probability

- So far we have studied *descriptive statistics*: methods of describing or summarizing a *sample*
- We want to move ahead to *inferential statistics*: methods for using the data in a sample to draw conclusions about the *population* from which the sample is drawn.
- Methods of inferential statistics are based on the question “How often would this method give a correct answer if I used it very, very many times?”
- The laws of probability relate to this question.

- A **statistic** is a numeric value that can be computed directly from sample data.
 - Example: we draw a sample of 10 Chinese males aged 21-65 years and measure the percent body fat of each one.
 - * The sample mean \bar{x} of the 10 data values is a statistic.
 - We do *not* need to use unknown parameters to compute a statistic.
 - We often use a statistic to *estimate* and unknown parameter.
 - But the exact value of a particular statistic will be different in different samples drawn from the same population.
 - * **sampling variability**

Randomness

- Chance behavior is unpredictable in the short run but has a predictable pattern in the long run.
- Example: tossing a coin
 - The proportion of heads in a small number of coin tosses is very variable.
 - As more and more tosses are done, the proportion settles down. It gets close to 0.5 and stays there.

Randomness

An experiment or observation is called **random** if individual outcomes are uncertain but there is a regular distribution of outcomes in a large number of independent repetitions.

Examples:

- We flip a coin and record the outcome as a head or tail
- We draw an 18-year-old American male “at random” and follow up to find out whether he lives to be 65
- We draw an American child at random and record his/her position in birth order of children in the family
- A researcher feeds a baby rat a particular diet and records the rat’s weight gain from birth to age 30 days

- French naturalist Count Buffon (1707-1788) tossed a coin 4040 times and got 2048 heads.
 - proportion heads: $\frac{2048}{4040} = 0.5069$
- While imprisoned by the Germans during World War II, South African mathematician John Kerrich tossed a coin 10,000 times and got 5067 heads.
 - proportion heads: $\frac{5067}{10,000} = 0.5067$
- In 1900 English statistician Karl Pearson tossed a coin 24,000 times and got 12,012 heads.
 - proportion heads: $\frac{12012}{24000} = 0.5005$
- American statistician Kate Cowles (19?? - 20??) tossed a coin 5 times and got 4 heads
 - proportion heads: $\frac{4}{5} = 0.8$
- She repeated the experiment and got 2 heads
 - proportion heads: $\frac{2}{5} = 0.4$

The **sample space S** is the set of all possible outcomes of a random experiment.

Examples:

- We flip a coin and record the outcome as a head or tail
- We draw an 18-year-old American male “at random” and follow up to find out whether he lives to be 65
- We draw an American child at random and record birth order
- A researcher feeds a baby rat a particular diet and records the rat’s weight gain from birth to age 30 days

An **event** is any outcome or set of outcomes of a random experiment.

Example: At random, we draw a child born in the US and record his/her live birth order. We would observe one of the following events:

She is

- 1st child
- 2nd child
- 3rd child
- 4th child
- 5th child
- 6th or later

Or, we might lump certain outcomes together into a single event of interest.

- Child is “1st child” or “not 1st child”

The **probability** of an event is the proportion of times the outcome would occur in a very long series of repetitions under the same conditions.

- (This is the “long-run frequency” definition of probability.)
- coin tosses: the probability of getting a head is 0.5
- birth order of randomly drawn American child

Birth order	1st	2nd	3rd	4th	5th	6+
Probability	0.416	0.330	0.158	0.058	0.021	0.017

The probability that an event occurs is often denoted with the letter P .

- $P(A)$ is the probability of event A

Capital letters near the beginning of the alphabet often are used to denote events.

Example:

- A might represent the event that the child is a 1st child.
- B might represent the event that the child is not a first child.

More probability terminology

- The event (A does not occur) is called the *complement of A* and represented by A^c
 - If A is the event that the randomly drawn child is a first-born child, then what is A^c ?
- Two events A and B that cannot occur simultaneously are *disjoint* or *mutually exclusive*
- The *union* of two events is the event that one or the other or both occur.
 - The union of events A and B is the event (A or B or both)
- The *intersection* of two events is the event that both occur.