Lecture 2

STAT 109: Introductory Biostatistics

Lecture 2: Probability as Data vs Theory

Learning Outcomes:

By the end of today's lecture, you should be able to:

1. Compute counts, proportions, and percentages of a particular event occurring in a number of trials.
2. Use symbols $(x, n, \hat{p}, P(\text{event}))$ correctly to describe observed data and theoretical probability.
3. Distinguish between observed outcomes (data) and theoretical probability.
4. Describe three rules that probabilities must satisfy.
5. State the definition of probability and connect it to the idea of long-run relative frequency in a repeatable random process.

Physical Randomness: Tossing a Coin

Let's toss a coin a few times in class, keeping track of the tosses that land "heads."

Rosanna's data

- Count of heads (x): _____
- Total tosses (n): _____
- Proportion ($\hat{p}$): _____
- Percentage: _____

Your data

- Count of heads (x): _____
- Total tosses (n): _____
- Proportion ($\hat{p}$): _____
- Percentage: _____

Your neighbor's data

Your neighbor, _____'s data:

- Count of heads (x): _____
- Total tosses (n): _____
- Proportion ($\hat{p}$): _____
- Percentage: _____

What's the highest count of heads? _____

Most tosses? _____

Highest proportion of heads? _____

Highest percentage of heads? _____

Comments:

If we care about comparing coins, then _____ is the best summary.

If we care about getting lots of "heads" then _____ is the best summary.

The summary of total tosses (n) tells us _____.

Question: Rosanna keeps tossing her coin forever. What would we expect?

- Data (what happened): _____
- Theory (long run): _____

Key definitions related to a trial

Fill in each definition in your own words:

- **Sample space:**

- **Outcome:**

- **Event:**

Notes on notation

Symbols and notation

We'll use two colors: one for observed data and another for theoretical quantities.

Class action required!

What color should Rosanna use for theoretical quantities that we imagine? _____

What color should Rosanna use for quantities that we can actually observe? _____

- x = count of "successes" in n observed trials
- n = number of observed trials
- $\hat{p} = x/n$ = observed proportion of successes in our trials
- Percentage of observed trials that were a success = $100 \cdot \hat{p}$
- $P(\text{event})$ = probability of an event

The notation for probability, $P(\text{event})$, is in the same format as notation for functions in high school algebra:

- = reads as _____
- () reads as _____ (not multiplication!)

$Y = f(x)$ reads as "y is the value of the function whose name is f when x is put in."

$Y = P(A)$ reads as "y is the value of probability when the event under consideration is A."

Example 1: Coin tosses

Suppose we toss a coin 20 times and see 15 heads.

- $x = 15$
- $n = 20$
- $\hat{p} = 15/20 = 0.75$
- Percentage $= 75\%$
- Theoretical $P(\text{Heads}) = 0.5$ (for a fair coin)

Student exercise 1: Coin tosses

A coin is tossed 10 times and comes up heads 3 times. Fill in:

- $x =$ _____
- $n =$ _____
- $\hat{p} =$ _____
- Percentage $=$ _____ %
- Theoretical $P(\text{Heads}) =$ _____

Example 2: Die rolls

Suppose a fair 6-sided die is rolled 24 times and we see the number "5" exactly 2 times.

- $x =$ _____
- $n =$ _____
- $\hat{p} =$ _____
- Percentage $=$ _____
- Theoretical $P("5") =$ _____

Student exercise 2: Die rolls

A die is rolled 30 times and shows "1" exactly 10 times. Fill in:

- $x =$ _____
- $n =$ _____
- $\hat{p} =$ _____
- Percentage $=$ _____ %
- Theoretical $P("1") =$ _____

Three rules for probabilities

All probabilities must satisfy:

1. $0 \leq P(A) \leq 1$ for any event A
2. Total probability across all possible outcomes $= 1$
3. If A and B are two events with no overlap, then $P(A \text{ or } B) = P(A) + P(B)$

Examples

- Coin: $P(\text{Heads}) + P(\text{Tails}) = $ _____
- Die: $P(\text{"1"}) + P(\text{"2"}) + \cdots + P(\text{"6"}) = $ _____
- Die: $P(\text{"1"}) + P(\text{"2"}) = $ _____
- $P(A) = 70$ is _____
- $-0.5 = P(B)$ is _____
- $P(D) = 0.7$ is _____

Definition of probability

$P(\text{Event}) = $ the proportion of times the event occurs out of the number of trials as the trial is repeated infinitely many times.

Probability is tied to:

- A specific random process (repeating a trial with uncertain outcome infinitely many times)
- Long-run relative frequency (a.k.a. "proportion") of an event occurring