

# Introduction to inference

- ▶ start with a question

Example: What proportion of adults in the United States approve of the job Barack Obama is doing as president?

Example: What is the mean annual precipitation (in inches) for Tacoma?

- ▶ each question is asked about a **population**

Example: all adults currently in the United States

Example: all years (assuming a constant climate)

- ▶ each question relates to a **variable** measured on the individual things in the population

Example: question response: approve or disapprove

Example: annual precipitation (in inches)

- ▶ each question is asked about a descriptor of the population distribution for the variable

Example: proportion of approve

Example: mean of annual precipitation

- ▶ a descriptor of interest for the population distribution of the variable is called a **parameter**

Example: proportion  $p$  of approve

Example: mean  $\mu$  of annual precipitation

- ▶ when population is too large or not accessible, measure the variable on a **sample**

Example: 1,042 people reached by telephone in the NY Times poll

Example: 47 years of precipitation data for a specific Tacoma weather station

- ▶ the corresponding descriptor of interest for the sample distribution of the variable is called a **statistic**

Example: proportion  $\hat{p} = 0.56$  of the 1,042 respondents who approve

Example: mean  $\bar{x} = 37.10$  inches for the 47 years of precipitation data

- ▶ the statistic is an **estimate** of the parameter
- ▶ to understand the nature of this estimate, need to understand the role that **randomness** plays in selecting a sample