Stat 217 – Day 3
More terminology

Previously:
- Observational unit: The entity of interest on which we collect data
- Variable: The characteristics we are measuring on the observational units
- Statistic: A number calculated from the observed data that summarizes the variable.
- Probability: The long-run proportion of times an outcome occurs if the random process is repeated indefinitely under identical conditions.

Example 1: The probability of winning anything in the Powerball lottery is .3014.

(a) Interpret this probability statement.

You expect to win \( \frac{3014}{10000} \) in many, many plays of the game.

(b) How do you think they determined this probability?

They looked at all possible outcomes and found the fraction of wins.

(c) How else might you estimate this probability?

Play thousands of times and calculate the proportion of wins.

Definitions:
- A simulation is an artificial recreation of a random process that allows us to investigate long-term properties of that process. Often we have to specify certain assumptions about the process.
- A parameter is a numerical summary of the underlying random process.

Example 2: What are the parameters in the Monty Hall game?

Probability of winning with stay strategy
Probability of winning with switch strategy

Example 3: An article published in *College Mathematics Journal* (Eyler, Shalla, Doumaux, and McDevitt, 2009) found that players of Rock-Paper-Scissors, particularly novices, tend to not prefer scissors. They had 119 people play the game against a computer, scissors was chosen \( \frac{1}{6} \)th of the time. Identify the observational units, variable, statistic, and parameter of interest.

Observational units: players

Variable: choice of scissors or not

Statistic: proportion of sample choosing scissors

Parameter: long-run probability of choosing scissors

Proposition: \( \frac{1}{6} \)
(b) What claim are these researchers making about the parameter?

\[
\text{less than } \frac{1}{3}
\]

Sometimes we can consider our observations to be a \textit{sample} for an ongoing \textit{process}. Other times our observations are coming from a \textit{sample} that is a subset of a larger \textit{population}. For example, with the Old Faithful eruptions (see text), there isn’t really a fixed population of eruptions out there for which we have measured a subset, but we may consider our data to be representative of the underlying eruption process. In both cases, we can define a parameter as a numerical summary of the variable in the underlying process or larger population.

\textbf{Example 4:} Would you consider the following data as arising from a population or a process? Identify the variable and parameter of interest.

(a) Can dogs sniff out cancer?

\begin{itemize}
  \item \text{Process:} dog deciding
  \item \text{Variable:} correct identification or not
  \item \text{Parameter:} long-run probability dog correct
\end{itemize}

(b) What proportion of American adults are optimistic about the new health-care laws?

\begin{itemize}
  \item \text{Population:} all American adults
  \item \text{Variable:} whether or not optimistic
  \item \text{Parameter:} proportion of all American adults who are optimistic
\end{itemize}

(c) Measurements are taken over a two-week period on the oxygen level in a river to assess it is in compliance with recommended levels.

\begin{itemize}
  \item \text{Process:} river
  \item \text{Variable:} measurements of oxygen
  \item \text{Parameter:} long-run probability of compliance
\end{itemize}

(d) How much money do college students tend to spend on textbooks in one semester/quarter?

\begin{itemize}
  \item \text{Population:} all college students
  \item \text{Variable:} amount spend
  \item \text{Parameter:} average spent by all college students
\end{itemize}

\textbf{To turn in:} The generation of children raised in the 1990s and 2000s have been dubbed “Generation M” because of the impact of media on their lives. The Kaiser Family Foundation conducted an extensive study to quantify how much time teenagers spend with various types of media. They gathered data from a nationally representative sample of 2032 teenagers, in particular asking whether or not the teen had a television in his or her bedroom. Identify the population or process of interest, the sample, the variable, the parameter of interest, and the statistic.