Is the dataset good enough for inference?

Read about some or most of these in the recommended text or one of the other texts. (Usually in Chapter 1)

1. Population versus sample

2. Observational study versus experimental study

3. Association between two variables

4. Explanatory variable versus response variable

5. Confounding variable
   One that is associated with BOTH the explanatory variable and response variable

6. Sampling methods
   Elementary statistics books only USE simple random sampling, but some of them (eg. Diaz) describe more complex sampling schemes such as stratified random sampling or cluster sampling.

7. Experimental design
   Random assignment of subjects to treatments evens out the differences in the effects of the possible confounding variables, so it is possible to obtain evidence for causation.

   Other ideas of good experimental design include control of outside variables, adequate replication, blocking, use of placebos, blind and double-blind experiments, matched pairs design.

8. What is needed for statistical inference?

   a. Random selection from the population allows for inference to the population

   b. In an experiment, random assignment of the values of the explanatory variable to the subjects supports inference about causation. (Rationale: It "breaks" any association between the explanatory variable and any possible confounding variable, because the results of the effect on the explanatory variable are randomly distributed among the treatments.)