

 Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > Producing Data: Designing Studies > Producing Data Summary > Summary (Producing Data)

 Bookmark this page

## Summary (Producing Data)

In this section, we discussed the first step in the big picture of statistics—**production of data**.

Production of data happens in two stages: **sampling** and **study design**.

Our goal in sampling is to get a **sample that represents the population of interest well**, so that when we get to the inference stage, making conclusions based on this sample about the entire population will make sense. We discussed several biased sampling plans, but also introduced the "family" of probability sampling plans, the simplest of which is the **simple random sample**, that (at least in theory) are supposed to provide a sample that is not subject to any biases.

In the section on study design, we introduced 3 types of design: observational study, controlled experiment, and sample survey.

We saw that with **observational studies** it is **difficult to establish** convincing evidence of a **causal relationship**, because of lack of control over outside variables (called lurking variables). Other pitfalls that may arise are that individuals' behaviors may be affected if they know they are participating in an observational study, and that individuals' memories may be faulty if they are asked to recall information from the past.

**Experiments** allow researchers to take control of lurking variables by **randomized assignment to treatments**, which helps provide more convincing evidence of causation. The design may be enhanced by making sure that subjects and/or researchers are **blind** to who receives what treatment. Depending on what relationship is being researched, it may be difficult to design an experiment whose setting is realistic enough that we can safely generalize the conclusions to real life.

Another reason that observational studies are utilized rather than experiments is that certain explanatory variables—such as income or alcohol intake—either cannot or should not be controlled by researchers.

**Sample surveys** are occasionally used to examine relationships, but often they assess values of many separate variables, such as respondents' **opinions** on various matters. Survey questions should be designed carefully, in order to ensure unbiased assessment of the variables' values.

Throughout this section, we established guidelines for the ideal production of data, which should be held as standards to strive for. Realistically, however, it is rarely possible to carry out a study which is completely free of flaws. Therefore, common sense must frequently be applied in order to decide which imperfections we can live with, and which ones could completely undermine a study's results.

---

Open Learning Initiative [🔗](#)



[🔗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [🔗](#).

© All Rights Reserved