Glossary

cluster sample A random sample obtained from a sampling scheme in which the population is divided into non-overlapping groups (usually based on geography) called clusters. Clusters are chosen using simple random sampling and each unit within the chosen cluster is measured.

confounding factor A factor that might provide an alternative explanation for any differences that appear between groups.

datacamp An online interactive environment for learning R.

explanatory (or treatment, or predictor) The variable on which the population is divided into groups. Note there can be finitely many or infinitely many such groups. Often this variable is explicitly set by the person designing the experiment or the statistician conducting an observational study.

observational study A method of collecting data in which researchers are unable to assign treatments, and so instead they observe some system as it is.

placebo A treatment in which nothing is applied to the experimental units.

placebo-controlled trial An experiment in which there is a placebo group.

protocol The way in which data is obtained from each of the sampled units.

random sample A sample from the population of interest, obtained using some chance mechanism.

randomized experiment A method of collecting data in which researchers are able to assign two or more "treatments" or "interventions" to subjects or units in their study, and in which some chance mechanism is used to make that assignment.

representative sample A sample in which the units have roughly the same characteristics as the units of the population.

R Studio An environment for running R.

response (or outcome) The variable being measured.

sample A subset of units from some population of interest.

simple random sample A random sample obtained from a sampling scheme in which every unit in the population has the same chance of being selected.

stratified sample A random sample obtained from a sampling scheme in which the population is divided into non-overlapping sub-populations and within each sub-population - or stratum - units are chosen using simple random sampling.

volunteer/convenience sample A non-random sample from the population of interest.