LING82100: experimental design and random variables

Kyle Gorman

1 Introduction

Linguistics is in part an *empirical science*: it relies on data to test hypotheses and develop new theories. A hypothesis is testable if we are able to collect data which bears on its predictions. Some examples of testable hypotheses in linguistics include:

- In Marathi, voiceless velar stops have longer average *voice onset times* (VOT) than coronal or labial stops.
- In English, word-final ⟨ing⟩ can be realized as either [ɪŋ] or [ɪn]. Apical realizations of this suffix are more common when it is the progressive suffix *-ing* than when it is part of a stem.
- In English, many verbs allow both prepositional indirect objects (e.g., *gave milk to the cat*) and double objects *gave the cat milk*. The double object construction is used more often when the recipient is pronominal than when it is a full noun phrase.

2 Experimental design

2.1 True experiments

When the researcher is able to control a variable of interest, the design is referred to as a (true) *experiment*.

- The variable or variables manipulated by the researcher is known as an *independent variable* (IV).
- The variable measured by the researcher is known as the *dependent variable* (DV).

The term *dependent* reflects our hypothesis that the DV "depends" on the IV and varies as some function of the IV. There are often multiple IVs manipulated in an experiment. While an actual experimental procedure can also measure multiple DVs, for statistical purposes we usually treat each DV as being part of a separate "experiment".

2.2 Quasi-experiments

When the IV(s) of interest are not under experimental control but are based on pre-existing differences, the design is referred to as a *quasi-experiment*. For instance, if one is studying language in children with or without autism spectrum disorder (ASD), it is not possible—nor would it be ethical to do so, were it possible—to "induce" ASD in some experimental subjects. Rather, the researcher must recruit two separate samples: children with ASD and children with typical development. However, the researcher must be careful not to interpret the influence of a quasi-IV on the DV without also controlling for other *confounding variables* (themselves quasi-IVs). For instance, children with ASD have on average lower verbal IQ and executive function, which may itself account for their poorer performance on an expressive vocabulary task.

The IVs of interest in quasi-experiments necessary involve a *between-subjects* design. This is so-called because the IVs—and other confounding variables—are constant for each subject and determined by the aforementioned pre-existing condition. While most true experiments expose each subject to multiple values of the IV(s)—a so-called *within-subjects* design—between-subjects designs are also used with true experiments, when, for example, some of the experimental conditions rely on deception or the experiment is likely to show practice effects between conditions.

2.3 Observational studies

When the researcher simply measures the associations between two or more variables of interest without any explicit manipulation or sampling, the design is referred to as an *observational study*. In such designs, there is no meaningful distinction between IVs and DVs; the variables of interest are often called *predictors* and *outcomes*, respectively, instead.

3 Populations and samples

Scientific theories often make statements about the complete, possibly-infinite *populations* of like individuals or cases, such as

- all voiceless stops in Marathi,
- all speakers of Philadelphia English, or
- all English ditransitive verbs.

However, the vast majority of research requires us to study a subset of the population, termed the *sample*. There are many reasons to study a sample rather than the full population—indeed, the population may be infinite.

Since the sample is smaller than the population, its properties —its *statistics*—may differ from those of the population. The difference between the properties of the sample and the population are known as *sampling error*. *Parametric statistics* allow us to *generalize*, i.e., to estimate the properties of the population (and the sampling error) from the statistics of our sample.

4 Scales of measurement

You are likely familiar with the notion of a *variable* from algebra, as in expressions like $y = 3x + 4$. In such expressions, if the value of one variable is known, it is possible to determine the exact value of the other. In contrast, the value of a *random variable* is in part due to chance; it cannot be computed directly, it can only be *sampled*. Random variables are described by their *distribution*, a collection of all values of that variable in a sample. How we describe the distribution of a random variable depends on its type or *scale*.

4.1 The Stevens scales

Stevens (1946) proposes a well-known four-level taxonomy for scales of measurement.

Categorical scale *Categorical* (or *nominal*) random variables have a fixed set of names or labels as their values. Some examples include correct/incorrect answer, ethnicity, or bilingual status. There is no *natural order* among the values of a categorical variable. When the values of a categorical variable are $\{\text{true}, \text{false}\}$, it is sometimes called *binomial* (or *binary*, or *boolean*).

Ordinal scale *Ordinal* (or *rank*) variables have rank values. They differ from categorical variables in that they have a natural order. It is important to note that we make no assumption about the size of the difference between ranks. For example, imagine that we rank all students in the class by their height, in descending order; then, the difference in height between the tallest and second-tallest student is not necessarily the same as the difference in height between the second-tallest and third-tallest student.

Interval scale *Interval* variables have values drawn from a numerical (integral or real-value) scale. Some examples include degrees Fahrenheit, dates, or scores on standardized tests. They differ from ordinal variables in that the difference between values is meaningful; for instance, if it was 60° F yesterday and 40° F today, we can sensibly say that it is 20° F warmer than yesterday.

Ratio scale *Ratio* variables differ from interval variables in that they have a unique, and non-arbitrary zero value, which gives a meaningful interpretation to ratios. Most physical measurements, including voice onset time or response times, are ratio scale. For instance, temperatures Kelvin are ratio scale because 0 K represents absolute zero, the absence of molecular motion. Thus it does make sense to say that 260° K is “twice” as hot as 130° K, but similar statements with Fahrenheit (or Celsius) are nonsensical.

4.2 Caveats

It is important to note that variables do not come to us labeled with their scale; rather, researchers impose interpretations onto them. For example, Likert [laɪ.kɝt] scale ratings (“on a scale from 1-7…”) are (minimally) ordinal variables, but many researchers treat them as interval variables for the purposes of statistical analysis.

4.3 Alternative scales

It is also important to note that the Stevens scales are not the only possible taxonomy. Mosteller and Tukey (1977), for instance, distinguish between

- *counted fractions*: fractions derived from counts, ranging between $[0, 1]$, and
- *counts*: non-negative integers.

They also distinguish between two types of ratios,

- *amounts*: non-negative real numbers, and
- *balances*: real numbers.

References

Mosteller, Frederick, and John W. Tukey. 1977. *Data analysis and regression: a second course in statistics.* Reading, MA: Addison-Wesley.

Stevens, Stanley S. 1946. On the theory of scales of measurement. *Science* 103:677–680.