

4.10: Formulas

The last kind of variable that I want to introduce before finally being able to start talking about statistics is the **formula**. Formulas were originally introduced into R as a convenient way to specify a particular type of statistical model (see Chapter15) but they're such handy things that they've spread. Formulas are now used in a lot of different contexts, so it makes sense to introduce them early.

Stated simply, a formula object is a variable, but it's a special type of variable that specifies a relationship between other variables. A formula is specified using the "tilde operator" `#~#`. A very simple example of a formula is shown below:⁶²

```
formula1 <- out ~ pred
formula1
```

```
## out ~ pred
```

The *precise* meaning of this formula depends on exactly what you want to do with it, but in broad terms it means "the `out` (outcome) variable, analysed in terms of the `pred` (predictor) variable". That said, although the simplest and most common form of a formula uses the "one variable on the left, one variable on the right" format, there are others. For instance, the following examples are all reasonably common

```
formula2 <- out ~ pred1 + pred2  # more than one variable on the right
formula3 <- out ~ pred1 * pred2  # different relationship between predictors
formula4 <- ~ var1 + var2        # a 'one-sided' formula
```

and there are many more variants besides. Formulas are pretty flexible things, and so different functions will make use of different formats, depending on what the function is intended to do.

This page titled [4.10: Formulas](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Danielle Navarro](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.