4 Introduction to regressio analysis

true

2022-03-22

Contents


```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Regression

Regression is a fundamental technique for predicting an outcome variable from one or more predictors. For instance, the average pitch of a speaker (the outcome) based on their gender (predictor 1) and the context of interaction (predictor 2); whether they will use a pronoun to refer to an entity (outcome) based on how predictable that entity is (predictor); or whether a linguistic school reform (predictor) increases the fluency of students in a language.

As summarized in Gelman et al. (**ros?**), some important uses of regression are: - prediction - exploring associations - extrapolation - causal inference

Linear regression

The simplest regression model relates predictors and the outcome in a linear fashion: a linear regression model. With x as a single predictor of y, we can write the model as:

$$y = \beta_0 + \beta_1 x + \text{error}$$

In words, outcome y is modelled as a linear combination of two parameters, β_0 and β_1, and an error term. The contribution of β_1 to the prediction is relative to the value of x, the predictor.

Let us explore these ideas with a concrete example (**franke+roettger:20?**), with study predicting voice pitch (outcome / y) based on whether the speaker is male or female (x). We will add further predictors later.

Case study: pitch

References

Corrections

If you spot a mistake or have suggestions, please get in touch with Thomas Brochhagen or create an issue