University of Essex

Online

Factors

It is commonly seen that we have to give numeric labels to categorical data. For example, if we have a sample of 100 people comprising 50 males and 50 females, it would be tedious to type "male" and "female" every time.

Alternatively, if "male" is coded as 1 and "female" is coded as 2, data entry becomes easier.

- The data structure available for such an arrangement is called a **factor**.
- Therefore, first, a vector needs to be created and then it needs to be transformed into a factor.
- For example, a factor with 10 elements can be created by executing the following two commands:

  > vec1<-c(1,2,2,1,1,2,1,2,1,2)

  > fac1<-factor(vec1)

- When *fac1* is typed in the command, it shows the following options:

  [1] 1 2 2 1 1 2 1 2 1 2

  Levels: 1 2

- The first line of output appears the same as that of a vector and the second line specifies that these are not numbers but 1 and 2 are levels and labelled as Male and Female respectively:

  > levels(fac1)=c("Male", "Female")

  > print(fac1)

  [1] Male Female Female Male Male Female Male Female Male Female

  Levels: Male Female