FACTORS

Factors

- Factors are used to represent categorical data. Factors can be unordered or ordered. They are useful for columns which have restricted value.
- R store each value of factor as a code internally. One can think of a factor as an integer vector where each integer has a label.
- Using factors with labels is better than numeric values because they become self describing. Having a variable that has values “Male” and “Female” is better than 0, 1

Factors

- R can have ordered factors which might represent things that are ranked. So they have an order but they are not numerical for example in universities there are assistant professor, associate professor, professor, so they are categorical and ranked.

Factors

- A factor includes not only the value of the corresponding categorical variable, but also the different possible levels of that variable (even if they are not present in the data). The function factor creates a factor with the following options.
- factor(x, levels, labels = levels , ordered = is.ordered(x))
- > x is a vector of data
- > levels :A vector of the values x might take or unique set of values of x. This is input value to factor.
- > labels: This is a character vector which is optional. It is used by the R to give output.
- > ordered: It keeps an order of the factor inputs and use it while presenting output.

> factor(c(1,1,1,2,2,2,3,3,3,4,4,4), labels=c(1,2,3), labels = c(“Q1”,”Q2”,”Q3”,”Q4”))

Factors in data frame

```
Height<-c(132,151,162,178,199)
Weight<-c(48,56,68,78,40)
Gender<-
c(“male”,”female”,”female”,”male”,”female”)
Data<-data.frame(height,weight,gender)
is.factor(data$gender) ## R converts the category data into factors by itself while creating data frame.
```

More on Factors

- Gl function generate the factors
- gl(n,k,labels=1:n,ordered=FALSE)
- N is an integer giving the number of levels
- K is an integer giving number of replications
- gl(12, 2, labels = c("Jan", "Feb","Mar","Apr","May","Jun","Jul","Aug","Sep", "Oct","Nov","Dec"))
- [1] Jan Jan Feb Feb Mar Mar Apr Apr May May Jun Jun Jul Jul Aug Aug Sep Sep Oct Oct Nov Nov Dec Dec

- Questions??