

Università degli Studi di Milano
Laurea Specialistica in Genomica Funzionale e Bioinformatica
Corso di Linguaggi di Programmazione per la Bioinformatica

Fattori

Giorgio Valentini
e-mail: valentini@dsi.unimi.it

DSI – Dipartimento di Scienze dell' Informazione
Università degli Studi di Milano

1

Fattori

- Strutture dati per rappresentare valori che possono assumere solo valori discreti
- In R i diversi valori che possono assumere i dati discreti sono definiti come *livelli*
- Es: dati qualitativi, o dati ordinali.
- Possono essere utili anche per “etichettare” dati memorizzati in altre strutture dati

2

Costruzione di fattori

I fattori sono costruiti tramite la funzione **factor**:

```
> trt <- factor( rep( c(" Control", "Treated"), c( 3, 4)))
> trt
[1] Control Control Control Treated Treated Treated Treated
Levels: Control Treated
> levels(trt) # visualizza i livelli del fattore
[1] " Control" "Treated"
> str( trt) # visualizza in modo succinto la struttura di
# un oggetto R
Factor w/ 2 levels "Control"," Treated": 1 1 1 2 2 2 2
> summary( trt) # la funzione summary fornisce una tabella
# della frequenze dei due livelli Control e Treated
Control Treated
3 4
```

I fattori si possono costruire con **factor** utilizzando un vettore esistente:

```
> s<-c(rep("A",3),rep("T",3),rep("G",3),rep("C",3))
> fs <- factor(s)
> fs
[1] A A A T T T G G G C C C
Levels: A C G T
> summary(fs)/length(fs)
A C G T
0.25 0.25 0.25 0.25
```

3

La funzione tapply

La funzione **tapply** consente di applicare una funzione f a dati etichettati da un fattore per ognuno dei livelli specificati nel fattore stesso:

```
> trt <- factor( rep( c(" Control", "Treated"), c( 3, 4)))
trt
[1] Control Control Control Treated Treated Treated Treated
Levels: Control Treated
> blood.press<-c(190,180,207,178,156,205,151) # vettore dati
# "etichettato" con trt
> tapply(blood.press,trt,mean) # la funzione mean è applicata ai
# dati blood.press per ogni livello di trt
Control Treated
192.3333 172.5000
```

4

Fattori ordinati

Un fattore ordinato è un tipo speciale di fattore in cui i livelli sono ordinati:

```
> pain <- ordered( c(" Moderate", "None", "Severe", "Severe", "None"),
+ levels = c(" None", "Moderate", "Severe"))
> str( pain)
Factor w/ 3 levels "None"," Moderate",...: 2 1 3 3 1
> pain
[1] Moderate None Severe Severe None
Levels: None < Moderate < Severe
> summary( pain)
None Moderate Severe
2 1 2
```

L'ordine di default è lessicografico (alfabetico):

```
> pain <- ordered( c(" Moderate", "None", "Severe", "Severe", "None"))
> pain
[1] Moderate None Severe Severe None
Levels: Moderate < None < Severe
```

5

Esercizi

1. Costruisci un fattore di 25 elementi (stringhe), caratterizzato da 3 diversi livelli
2. Costruisci un fattore ordinato di 15 elementi di livelli basso < medio < alto. Genera un corrispondente vettore di misurazioni di lunghezza 15 e calcola il valor medio e la deviazione standard (funzione sd) di ogni livello tramite la funzione tapply.
3. Quali strutture dati si potrebbero scegliere per modellare un data set caratterizzato da un insieme di pazienti sani e malati sottoposti a 5 diversi tipi di analisi cliniche?

6