

Ricodificare le variabili

La ricodifica consiste nella trasformazione di una variabile in modo che le stesse informazioni vengano espresse attraverso codici diversi rispetto a quelli originari.

Rinominare

Vedi

- [Rinominare le variabili](#)
- [select, rename \(dplyr\)](#)

Trasformare

Da vettore a fattore, e viceversa; gestire i formati delle date.

Vedi [Trasformare il tipo di variabile](#)

Trasformare i livelli di un fattore

Vedi:

- [recode, recode_factor](#): sostituire i valori
- [fct_recode](#): cambiare i livelli dei fattori
- [fct_lump](#): ricodificare le modalità meno frequenti in “altro”

Eliminare i livelli non usati

Se, ad esempio dopo una ricodifica o un subset dei dati, possono rimanere livelli non utilizzati, ovvero con frequenza pari a 0, usiamo la funzione `droplevels`:

```
survey$Smoke2 <- droplevels(survey$Smoke2)
```

Ricodificare: modificare i valori

Vedi [mutate, transmute \(dplyr\)](#)

In linea di massima si tratta di **assegnare** certi valori ad una variabile preesistente, oppure di costruire una nuova variabile, assegnandoi valori ai casi, in base a certe condizioni.

Il comando di base è dunque quello di **assegnamento**, nella forma `valori di X ← nuovi valori di X`.

Per recuperare i casi che presentano un certo valore di una variabile X (vedi la voce **dataframe** e **indicizzazione**), si usa l'espressione:

```
X[X==1]  
X[X=="donna"]
```

Per individuare i valori, si possono naturalmente utilizzare tutti i consueti **Operatori**.

Ai casi così identificati, si assegna un nuovo valore:

```
X[X=="donna"] <- "D"  
  
# o anche, per la stessa variabile  
X[X=="donna"] = "D"
```

Variabili numeriche

Calcolare una nuova variabile

Nel dataset **Istat16.rda** (scaricabile), creiamo la variabile Totale (della popolazione), sommando Totale.Maschi e Totale.Femmine:

```
Istat16$Totale <- Istat16$Totale.Maschi + Istat16$Totale.Femmine
```

Naturalmente, possiamo utilizzare espressioni di trasformazione più complesse.

Standardizzare e normalizzare

Vedi **Standardizzazione delle variabili**

Variabili categoriali (Fattori)

Si vuole dicotomizzare in Sì/No, la variabile Smoke del dataset “survey” (del pacchetto MASS).

La procedura è sostanzialmente la stessa illustrata precedentemente:

```
# lavoro su una copia di survey
require(MASS)
mysurvey <- survey

# creo la nuova variabile
mysurvey$Smoke2[mysurvey$Smoke=="Heavy"] <- "Si"
mysurvey$Smoke2[mysurvey$Smoke=="Occas"] <- "Si"
mysurvey$Smoke2[mysurvey$Smoke=="Regul"] <- "Si"
mysurvey$Smoke2[mysurvey$Smoke=="Never"] <- "No"
```

Variabili di testo

Vedi: [Modificare le stringhe e i vettori carattere](#)

Aggregare in classi

Vedi:

- [recode](#), [recode_factor](#): sostituire i valori
- [fct_lump](#): ricodificare le modalità meno frequenti in “altro”

Si vogliono aggregare i valori della variabile speed del dataset “cars”, in tre livelli: bassa, media, alta. Si tratta di costruire una nuova variabile ordinale (speed2) e trasformarla in fattore ordinato.

```
# lavoro su una copia di cars
mycars <- cars

# ricodifico in una nuova variabile
mycars$speed2[mycars$speed<=13] = "bassa"
mycars$speed2[mycars$speed>13 & mycars$speed<=17] = "media"
mycars$speed2[mycars$speed>17] = "alta"

# trasformo la nuova variabile in un fattore ordinato
mycars$speed2 <- factor(mycars$speed2)
mycars$speed2 <- ordered(mycars$speed2,
                        levels=c("bassa", "media", "alta"))
```

recode

Vedi [recode \(car\)](#)

La ricodifica con RCommander

Vedi [La ricodifica delle variabili con RCommander](#)

Script di esempio

[Es Ricod 2.R](#)

```
# LIVELLI DI UN FATTORE

# dati
library(MASS)
data(survey)

# rinomino il livello del fattore
levels(survey$Smoke2)[levels(survey$Smoke2)=="Never"] <- "Mai"

# elimino i livelli non usati
survey$Smoke2 <- droplevels(survey$Smoke2)

# VARIABILE CARDINALE

# calcolare una nuova variabile
Istat16$Totale <- Istat16$Totale.Maschi + Istat16$Totale.Femmine

# aggregare in classi

# lavoro su una copia di cars
mycars <- cars

# ricodifico in una nuova variabile
mycars$speed2[mycars$speed<=13] = "bassa"
mycars$speed2[mycars$speed>13 & mycars$speed<=17] = "media"
mycars$speed2[mycars$speed>17] = "alta"

# trasformo la nuova variabile in un fattore ordinato
mycars$speed2<-factor(mycars$speed2)
```

```
mycars$speed2 <- ordered(mycars$speed2,  
                          levels=c("bassa", "media", "alta"))  
  
# VARIABILE CATEGORIALE  
  
# lavoro su una copia di survey  
mysurvey <- survey  
  
# ricodifico in una nuova variabile  
mysurvey$Smoke2[mysurvey$Smoke=="Heavy"] <- "Si"  
mysurvey$Smoke2[mysurvey$Smoke=="Occas"] <- "Si"  
mysurvey$Smoke2[mysurvey$Smoke=="Regul"] <- "Si"  
mysurvey$Smoke2[mysurvey$Smoke=="Never"] <- "No"
```

Gestione dei dati

From:

<https://www.agnesevardanega.eu/wiki/> - Ricerca Sociale con R

Permanent link:

https://www.agnesevardanega.eu/wiki/r/gestione_dei_dati/ricodifica_delle_variabili

Last update: **13/08/2025 10:53**

