

Attività di allineamento

Laboratorio Informatico-Statistico



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Corso di Laurea Magistrale in Statistica Economia e Impresa

Dipartimento di Scienze Statistiche Paolo Fortunati

Anno Accademico 2021/2022



Next steps

Next steps

For loop

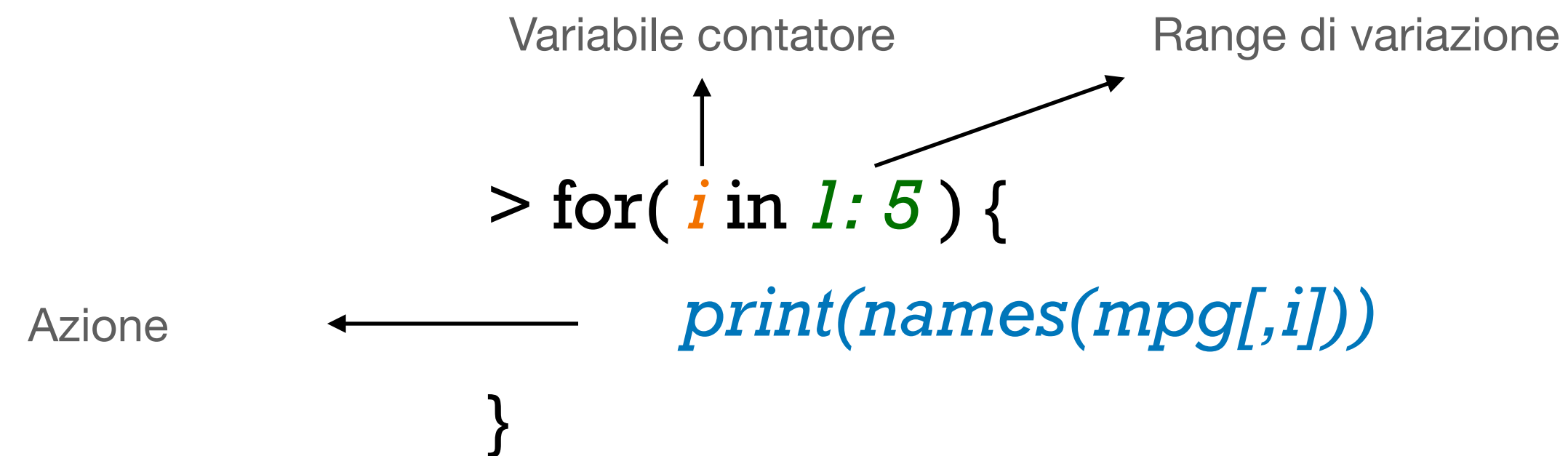
Utile quando si deve eseguire un azione su più elementi

```
> for(variabile contatore in range di variazione) {  
    azione  
}
```

Next steps

For loop

Utilizzando il df *mpg*, per stampare il nome delle prime 5 variabili:



The diagram shows a for loop with three annotations: 'Variabile contatore' pointing to the variable 'i', 'Range di variazione' pointing to the range '1:5', and 'Azione' pointing to the print statement.

```
> for(i in 1:5) {  
    print(names(mpg[,i]))  
}
```

n.b. questo esempio serve ad apprendere il funzionamento del ciclo, quale sarebbe stato il modo più efficiente per vedere il nome delle prime 5 colonne?

```
> names(mpg)[1:5]
```

Next steps

If - else

Dato l'oggetto $a = 5$, la condizione

```
> if ( is.numeric(a) ) {  
    print(" l'oggetto a è un numero ")  
} else {  
    print(" l'oggetto a non è un numero ")  
}
```

Restituirà “ l'oggetto a è un numero “ in quanto la condizione risulta essere soddisfatta

Next steps

For loop & If - else

Combiniamo il ciclo for e la condizione if - else in modo da estrarre degli indici sintetici (mediana, range interquartile) dalle variabili del df mpg nel caso in cui siano numeriche

```
>variables = names(mpg)
>out = c()
> for (i in 1:length(variables)) {
  if (is.numeric(mpg[, variables[i]])){
    out = c(out, paste("mediana :", as.numeric(summary(mpg$displ)[3]))) )
  } else {
    out = c(out, paste("(non numerica)"))
  }
}
```

Next steps

Functions

Utile quando per snellire il codice e renderlo più ordinato

```
> nomedellafunzione = function(argomenti) {  
    azioni  
    return(oggettodarestituire)  
}
```

Gli *argomenti* possono essere obbligatori o facoltativi (o anche assenti). Un argomento facoltativo viene accompagnato da un valore di default, uno obbligatorio no.

Next steps

Exercises

1. Importare come `data.frame` il dataset `airquality` del pacchetto *datasets*
2. Contare i null values per colonna
3. Eliminare i null values
4. Trasformare la colonna *Temp* (espressa in °F) in °C
5. Trasformare la colonna *Wind* (espressa in miglia/h) in km/h
6. Qual'è la correlazione fra il livello di ozono e la temperatura?
7. Produrre un grafico con la temperatura media per mese

Next steps

Portali utili

1. github.com
2. [kaggle.com](https://www.kaggle.com)
3. towardsdatascience.com
4. stackoverflow.com

Next steps

Reach me



raffaele.anselmo2@unibo.it: *domande sul corso*

r.anselmo@crif.com : *curiosità sul mondo professionale in ambito data science*



<https://www.linkedin.com/in/raffaele-anselmo/>



<https://github.com/RaffaeleAns>

Thanks

Raffaella Salvani