

# Dall'Analisi Fattoriale alla Regressione Lineare

*Metodi Quantitativi per Economia,  
Finanza e Management*

*Esercitazione n° 10*

# Consegna Lavoro di gruppo

- La scadenza per la consegna del lavoro di gruppo è fissata inderogabilmente per il giorno:

**Venerdì 13 Gennaio 2017**

- La consegna va effettuata **entro le ore 12** alla **Sig.ra Enrica Luezza** (Segreteria 4° Piano)
- Il materiale da consegnare consiste in:
  - stampa cartacea della presentazione in Power Point;
  - Chiavetta USB contenente:
    - questionario;
    - base dati in formato Excel;
    - presentazione Power Point

**N.B.** Il supporto elettronico (chiavetta USB) non sarà restituito.

# Processo dell' analisi fattoriale

Identificazione  $p$  variabili di partenza (variabili quantitative o scale di punteggio)



Selezione di  $k$  fattori  
(dove  $k < p$ )

Utilizzo di alcuni criteri per la *selezione dei possibili valori di  $k$*  (è possibile identificare più valori di  $k$  adeguati)

*Confronto* tra le possibili soluzioni identificate (confronto delle comunalità)

*Verifica dell'interpretabilità* della soluzione scelta ed eventuale indagine di una soluzione differente



Interpretazione della soluzione finale

# Analisi fattoriale (1/2)

STEP 1: scegliere quanti fattori considerare (scelta di varie soluzioni)

- la regola autovalori  $> 1$
- lettura dello SCREE PLOT
- Circa 1/3 delle variabili originarie
- Variabilità spiegata  $> 60\%$

```
Nome_1 = princomp(nome_subset, cor=TRUE)
```

```
get_eigenvalue(nome_1)
```

```
plot(nome_1, type='lines')
```

STEP 2: confrontare le soluzioni scelte

- cumunalità finali

```
Nome_2= principal(nome_subset, residuals=FALSE, nfactors = num_fattori,  
rotate = 'none')
```

```
Nome_2$communality
```



# Analisi fattoriale (2/2)

STEP 3: una volta scelta la soluzione finale

- ruotare i fattori
- interpretare i fattori
- salvare il data set con i fattori e rinominarli

```
nome_3= principal(nome_subset, residuals=FALSE, nfactors = num_fattori,  
rotate = 'varimax', score=TRUE)
```

```
nome_3$loadings
```

```
fa.diagram(nome_3)
```

```
Nome_4=cbind(dataset_originale, nome_2$scores)
```

STEP 4: se l'interpretazione non è soddisfacente ripetere lo step n°3 variando metodo di rotazione o provando un'altra soluzione.



# Regressione lineare (1/3)

1. Individuazione variabili dipendente e regressori
2. Trasformazione di eventuali variabili qualitative in dummy
3. Stimare un modello di regressione lineare utilizzando la procedura automatica di selezione delle variabili (stepwise)
4. Valutare la bontà del modello (R-square, Test F, Test t)
5. Se la procedura stepwise non ha prodotto tutte stime significative, provare a stimare un modello di regressione lineare con i soli parametri le cui stime sono significative. Tornare al punto 4, poi al punto 6.



# Regressione lineare (2/3)

6. Verificare la presenza di multicollinearità (se i regressori del modello sono i fattori di un'analisi fattoriale non è necessario perchè risultano non correlati per costruzione → tutti i  $VIF_j = 1$ )
  - ✓ Se si è in presenza di multicollinearità: azioni per eliminarla e ripetere i punti 3, 4
  - ✓ In assenza di multicollinearità: passare al punto 7
7. Verificare l'impatto dei regressori nella spiegazione del fenomeno (ordinarli usando il valore assoluto dei coefficienti standardizzati e controllare il segno dei coefficienti)
8. Interpretazione del coefficienti standardizzati



# Regressione lineare (3/3)

## Sintassi

- Stimare un modello di regressione lineare

```
Nome_dataset_lm = lm (variabile_dipendente ~  
variabile_indipendente, data=dataset_input)
```

- Stimare un modello di regressione lineare con il metodo stepwise

```
nome_modello_lm = step(nome_dataset_lm, direction='both')
```

- Calcolare le stime standardizzate dei coefficienti

```
lm.beta = (nome_modello_lm)
```

- Verifica presenza multicollinearità

```
vif(nome_subset_input)
```





# Esercizio

Il dataset `ct_telefonia.csv` contiene i dati di 126.761 clienti di una compagnia telefonica e 25 variabili quantitative.

#	Variable	Descrizione
1	AMMONT_RICARICA_BONUS	Ammontare ricariche bonus
2	AMMONT_RICARICA_PAG	Ammontare ricariche pagate dal cliente
3	AMMONT_RICARICA_PAG_LOTTO	Ammontare ricariche effettuate tramite circuito lotto sisal
4	AMMONT_RICARICA_RICORRENTE	Ammontare ricariche ricorrenti
5	ANZIANITA_SIM	Anzianità della SIM espressa in mesi
6	CONTATTI_INBOUND	Numero di volte in cui il cliente ha contattato il call center negli ultimi 6 mesi
7	CONTATTI_OUTBOUND	Numero di volte in cui il call center ha contattato il cliente negli ultimi 6 mesi (per campagna commerciale)
8	D_OPZ_ESTERO	Variabile che indica se è attiva, disattiva o dismessa l'opzione telefonate vantaggiose verso l'estero
9	D_OP_NUM_PREF	Variabile che indica se è attiva, disattiva o dismessa l'opzione telefonate vantaggiose verso un numero preferito
10	D_RIC_RICORRENTE	Variabile che indica se è attiva, disattiva o dismessa l'opzione di ricariche ricorrente
11	ETA_CUSTOMER	Età del cliente
12	FLAG_OPZ_COUNTRY	Flag che indica se è stato scelto un particolare paese per effettuare chiamate vantaggiose
13	GENDER	Genere
14	ID_CUSTOMER	ID Cliente
15	MINUTI_ASSISTENZA	Minuti chiamate effettuate dal cliente per ricevere assistenza dall'operatore, negli ultimi 6 mesi
16	MINUTI_VOCE_ITZ	Minuti voce verso direttrici internazionali negli ultimi 6 mesi
17	MINUTI_VOCE_OFFNET	Minuti voce offnet (SIM di altri operatori) negli ultimi 6 mesi
18	MINUTI_VOCE_ONNET	Minuti voce onnet (SIM dello stesso operatore) negli ultimi 6 mesi
19	NUMERO_RICARICHE_BONUS	Numero di ricariche bonus negli ultimi 6 mesi
20	NUMERO_RICARICHE_RICORRENTI	Numero di ricariche ricorrenti negli ultimi 6 mesi
21	RECENZA_RICARICA_BONUS	Mesi trascorsi dall'ultima volta in cui il cliente ha ricevuto una ricarica bonus
22	REC_CONT_INBOUND	Mesi trascorsi dall'ultima volta in cui il cliente ha contattato il call center
23	REC_CONT_OUTBOUND	Mesi trascorsi dall'ultima volta in cui il call center ha contattato il cliente
24	SIM_ATTIVE	Numero di SIM attive per cliente
25	ARPU	Valore arpu: ricavi medi ottenuti mensilmente per ciascun utente

# Esercizio

1. Cambiare la directory di lavoro in quella in cui si è salvato il dataset e salvare il dataset in un oggetto in R.
2. Effettuare un'analisi fattoriale utilizzando le seguenti variabili:

CONTATTI\_INBOUND  
CONTATTI\_OUTBOUND  
REC\_CONT\_INBOUND  
REC\_CONT\_OUTBOUND  
MINUTI\_ASSISTENZA  
MINUTI\_VOCE\_ITZ  
MINUTI\_VOCE\_OFFNET  
MINUTI\_VOCE\_ONNET  
RECENZA\_RICARICA\_BONUS  
AMMONT\_RICARICA\_BONUS  
AMMONT\_RICARICA\_PAG  
AMMONT\_RICARICA\_PAG\_LOTTO\_SISAL  
AMMONT\_RICARICA\_RICORRENTE  
NUMERO\_RICARICHE\_BONUS  
NUMERO\_RICARICHE\_RICORRENTI  
FLAG\_OPZ\_COUNTRY

# Esercizio

- Scegliere il numero di fattori ottimali
- Salvare i fattori interpretati in un nuovo dataset

## 3. Stimare un modello di regressione lineare utilizzando

- come variabile dipendente il valore dell'Arpu
- come potenziali regressori, oltre ai fattori individuati al punto precedente, anche le variabili: età del cliente, anzianità della sim e numero di sim attive per cliente:
- Utilizzare l'opzione di stepwise
- Effettuare tutti i passaggi presenti nelle slide di riepilogo, rispondendo anche alle seguenti domande:
  - a. Il valore dell'R-quadro è soddisfacente?
  - b. Cosa possiamo affermare osservando i dati relativi al test F e ai test t?
  - c. Quale regressore influenza maggiormente la variabile dipendente?