

Esercizi riepilogativi

Analisi Univariata e Bivariata

*Metodi Quantitativi per Economia,
Finanza e Management*

Esercitazione n°6

Metodi Quantitativi per Economia, Finanza e Management

Obiettivi di questa esercitazione:



Metodi Quantitativi per Economia, Finanza e Management

Tipo di analisi	Strumenti	Lavoro di gruppo
ANALISI UNIVARIATA	<ul style="list-style-type: none"> - DISTRIBUZIONI DI FREQUENZA - INDICI DI POSIZIONE (MISURE DI TENDENZA CENTRALE E MISURE DI TENDENZA NON CENTRALE) - INDICI DI DISPERSIONE - MISURE DI FORMA DELLA DISTRIBUZIONE 	OBBLIGATORIO
ANALISI BIVARIATA E TEST STATISTICI PER LO STUDIO DELL'ASSOCIAZIONE TRA VARIABILI	<p><u>Due variabili qualitative o quantitative discrete:</u> TABELLA DI CONTINGENZA E INDICI CHI QUADRO E V DI CRAMER TEST CHI QUADRO PER L'INDIPENDENZA STATISTICA</p> <p><u>Due variabili quantitative continue:</u> INDICE DI CORRELAZIONE DI PEARSON (ρ) E COVARIANZA TEST t PER L'INDIPENDENZA LINEARE</p> <p><u>Una variabile qualitativa e una quantitativa continua:</u> INDICE η^2 TEST F PER L'INDIPENDENZA IN MEDIA</p>	OBBLIGATORIO
ANALISI MULTIVARIATA	<ul style="list-style-type: none"> - ANALISI FATTORIALE - REGRESSIONE LINEARE - REGRESSIONE LOGISTICA 	OBBLIGATORIO

Analisi Univariata: Misure di Sintesi

Misure di posizione:

Misure di tendenza centrale:

- Media aritmetica
- Mediana
- Moda

Misure di tendenza non centrale:

- Quantili di ordine p (percentili, quartili)

Misure di dispersione:

- Campo di variazione
- Differenza interquartile
- Varianza
- Scarto quadratico medio
- Coefficiente di variazione

Misure di forma della distribuzione:

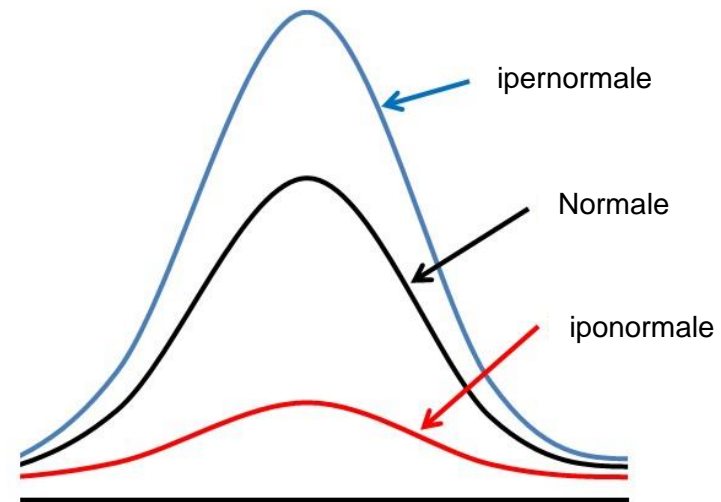
- Skewness
- Kurtosis



Misure di Forma della Distribuzione

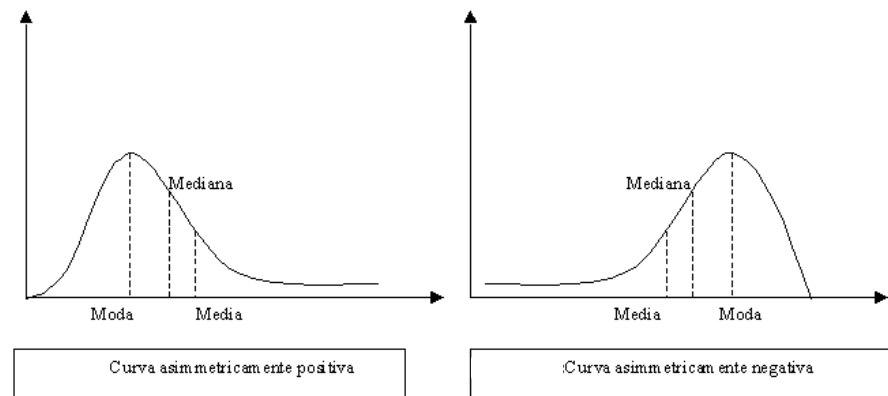
Kurtosis: indice che permette di verificare se i dati seguono una distribuzione di tipo Normale (simmetrica)

- $\beta=3$ se la distribuzione è “Normale”
- $\beta<3$ se la distribuzione è iponormale
- $\beta>3$ se la distribuzione è ipernormale



Skewness: indice che informa circa il grado di simmetria o asimmetria di una distribuzione

- $\gamma=0$ distribuzione simmetrica
- $\gamma<0$ asimmetria negativa (mediana > media)
- $\gamma>0$ asimmetria positiva (mediana < media)



Analisi Univariata: Procedure R

Studio della distribuzione di ogni variabile, singolarmente considerata, all'interno della popolazione

R	TIPO VARIABILE	FUNZIONE
Freq table frequencyBy	Variabili qualitative o quantitative discrete	Distribuzione di frequenze (frequenze assolute, relative)
summary basicStats IQR CV getmode	Variabili quantitative	Calcolo misure di sintesi di tipo univariato

NB: per utilizzare le funzioni frequencyBy e getmode è necessario lanciare il corpo delle funzioni prima di utilizzarle (esercitazione 3).



Analisi Bivariata: Riepilogo

Studio della distribuzione di due variabili congiuntamente considerate e delle relazioni esistenti tra esse

TIPO DI VARIABILI	TIPO DI RELAZIONE INDAGATA	INDICI DI DIPENDENZA	FUNZIONI R
Due Variabili Qualitative	Indipendenza Statistica	Chi Quadro, Cramer V	CrossTable, CramerV
Due Variabili Quantitative	Indipendenza Lineare	coeff. di correlazione lineare	cor.test
Una Qualitative e Una Quantitativa continua	Indipendenza in media	indice eta-quadro	etaSquared



Test d'Ipotesi

1) Sistema di Ipotesi



2) Calcolo Statistica test



3) Calcolo P-value

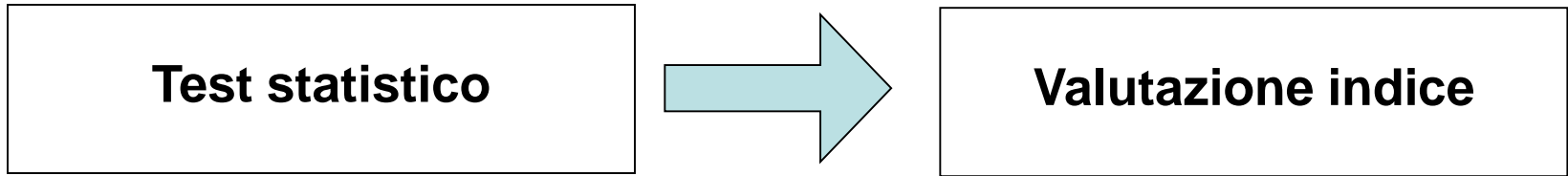


- Formulazione ipotesi nulla e ipotesi alternative
 - { H_0 : X e Y sono indipendenti
 - { H_1 : X e Y dipendenti
- Impostazione a priori del livello di significatività α
- Calcolo del valore della statistica test (specifica del test che si sta conducendo) sulla base dei dati campionari
- Calcolo del livello di significatività osservato

- Se **p-value** $< \alpha$ → sulla base dei dati campionari, la probabilità di rifiutare H_0 quando H_0 è vera è inferiore alla soglia scelta → **rifiuto H_0**
- Se **p-value** $\geq \alpha$ → **accetto H_0**



Approccio di analisi



- 1) Eseguire l'opportuno test statistico a seconda della tipologia delle variabili poste a confronto;
- 2) Analizzare l'esito del test (p-value confrontato con α)
- 3) Valuto l'indice opportuno:
 - a) Se il test mi indica indipendenza ($p\text{-value} > \alpha$) \rightarrow verifico se il valore dell'indice di dipendenza conferma l'esito del test;
 - b) Se il test mi indica dipendenza ($p\text{-value} < \alpha$) \rightarrow valuto il valore dell'indice di dipendenza per indagare la forza della relazione.



ATTENZIONE

R utilizza la notazione scientifica per riportare alcuni risultati che non riesce a stampare nel prompt dei comandi, poiché numeri o troppo grandi o troppo piccoli.

VALORE GRANDE:

1.197e+07, sposto la virgola a destra di 7 posizioni → 11 970 000

VALORE PICCOLO:

1.197e-04, sposto la virgola a sinistra di 4 posizioni → 0.0001197



Metodi Quantitativi per Economia, Finanza e Management

Obiettivi di questa esercitazione:



Dataset

Il dataset Banca contiene informazioni relative ai comportamenti di 1000 clienti di una banca.

Le variabili presenti nel dataset riguardano informazioni sullo stato economico del cliente (es. saldo CC, importi vari) e informazioni contrattuali (es. utilizzo plastic card, accredito stipendio).



Variabili (1)

#	Variable	Type	Len	Label
1	UTE	Num	8	IMPORTO TOTALE UTENZE DOMICILIATE
2	NUTE	Num	8	NUM. OPERAZIONI UTENZE DOMICILIATE
3	CARBCM	Num	8	IMPORTO TOTALE PLASTIC CARDS
4	NCARBCM	Num	8	NUM. OPERAZIONI PLASTIC CARDS
5	STICONV	Num	8	IMPORTO TOTALE ACCREDITO STIPENDIO
6	NSTICONV	Num	8	NUM. OPERAZIONI ACCREDITO STIPENDIO
7	TIT	Num	8	IMPORTO TOTALE CAUSALE TITOLI
8	NTIT	Num	8	NUM. OPERAZIONI CAUSALE TITOLI
9	RIMBAFF	Num	8	IMPORTO TOTALE RIMBORSI AFFIDAMENTI
10	NRIMBAFF	Num	8	NUM. OPERAZIONI RIMBORSI AFFIDAMENTI
11	LEAS	Num	8	IMPORTO TOTALE CAUSALE LEASING
12	NLEAS	Num	8	NUM. OPERAZIONI CAUSALE LEASING
13	ASS	Num	8	IMPORTO TOTALE ASSEGNI
14	NASS	Num	8	NUM. OPERAZIONI ASSEGNI
15	TASSE	Num	8	IMPORTO TOTALE CAUSALE TASSE
16	NTASSE	Num	8	NUM. OPERAZIONI CAUSALE TASSE
17	ASSIC	Num	8	IMPORTO TOTALE POLIZZE ASSICURATIVE
18	NASSIC	Num	8	NUM. OPERAZIONI POLIZZE ASSICURATIVE
19	CD	Num	8	IMPORTO TOTALE CERTIFICATI DI DEPOSITO
20	NCD	Num	8	NUM. OPERAZIONI CERTIFICATI DI DEPOSITO



Variabili (2)

#	Variable	Type	Len	Label
21	CUSTODIA	Num	8	IMPORTO COMMISSIONI TITOLI A CUSTODIA
22	NCUST	Num	8	NUM. OPER. COMMISSIONI TITOLI A CUSTODIA
23	IMPA	Num	8	IMPORTO OPERAZIONI AVERE
24	IMPAP	Num	8	IMPORTO OPERAZIONI AVERE ANNO PRECEDENTE
25	IMPD	Num	8	IMPORTO OPERAZIONI DARE
26	IMPDP	Num	8	IMPORTO OPERAZIONI DARE ANNO PRECEDENTE
27	NAVEP	Num	8	NUM. OPERAZIONI AVERE ANNO PRECEDENTE
28	NAVERE	Num	8	NUM. OPERAZIONI AVERE
29	NDARE	Num	8	NUM. OPERAZIONI DARE
30	NDARP	Num	8	NUM. OPERAZIONI DARE ANNO PRECEDENTE
31	SMEDIO	Num	8	SALDO MEDIO DI CONTO CORRENTE
32	S_UTENZE	Num	8	DOMICILIAZIONE UTENZE S/NO
33	S_PLASTI	Num	8	UTILIZZO PLASTIC CARD S/NO
34	S_STIPEN	Num	8	ACCREDITO AUTOMATICO STIPENDIO S/NO
35	S_TITOLI	Num	8	POSSESSO TITOLI S/NO
36	S_RIMBOR	Num	8	UTILIZZO AFFIDAMENTI
37	S_LEASIN	Num	8	LEASING S/NO
38	S_ASSICU	Num	8	POLIZZE ASSICURATIVE S/NO
39	S_CERTIF	Num	8	POSSESSO CERTIFICATI DI DEPOSITO
40	S_CUSTOD	Num	8	POSSESSO TITOLI A CUSTODIA
41	CLIENTE	Num	8	CODICE PROGRESSIVO CLIENTE
42	sezzo	Char	1	SESSO DEL CLIENTE



Modalità di svolgimento

- ✓ Impostare l'area di lavoro sulla cartella che contiene il dataset **banca.csv**
- ✓ Importare il file `banca.csv`, salvarlo in un oggetto a cui attribuire il nome **banca**
- ✓ Svolgere gli esercizi utilizzando il dataset **banca**.
- ✓ Riportare i risultati delle analisi svolte (copiando gli output R necessari e commentando adeguatamente) nel file `Esercitazione_6.docx` contenuto nella cartella `C:\corso`



Esercizi (1)

1. Qual è la media del saldo medio del conto corrente [SMEDIO]?
 2. Si può affermare che la maggior parte dei clienti utilizza la plastic card [S_PLASTI]?
 3. E' corretto affermare che il 66.67% dei clienti non possiede titoli [S_TITOLI]?
 4. E' corretto affermare che il saldo medio del conto corrente è una variabile con asimmetria positiva [SMEDIO]? Riportare anche il grafico.
- 4.Bis Qual è la moda del numero di operazioni delle utenze domiciliate?[NUTE]



Esercizi (2)

5. Tra coloro che hanno una polizza assicurativa, qual è l'importo medio delle polizze assicurative [ASSIC][S_ASSICU]?
6. Chi ha effettuato molte 'operazioni dare' l'anno precedente ha effettuato molte 'operazioni dare' nell'anno analizzato [NDARE][NDARP]?
- 6.Bis Esiste una relazione tra le variabili saldo medio in c/c, accredito stipendio, importo commissioni titoli e importo polizze assicurative? [SMEDIO][STICONV][CUSTODIA][ASSIC]
7. Si può affermare che esiste una relazione tra il numero di 'operazioni dare' e 'operazioni avere' [NDARE][NAVERE]?



Esercizi (3)

8. C'è una relazione tra la domiciliazione delle utenze e l'utilizzo della plastic card [S_UTENZE][S_PLASTI]? E tra il possesso di certificati di deposito e il possesso di titoli a custodia [S_CERTIF][S_CUSTOD]?
9. Si può affermare che esiste una relazione tra il sesso e l'accredito stipendio [SESSO][STICONV]?
10. Si può affermare che esiste una relazione tra possesso di titoli e saldo medio in conto corrente? [S_TITOLI] [SMEDIO]

