www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

STATISTCA 2018/2019

Programma:
- Nozioni introduttive (Parte 1)
- Distribuzione statistiche e rappresentazione i grafiche (Parte 2)
- Statistica univariata: (Parte 3, 4, 5,)
  - Media
  - Variabillità
  - Altre caratteristiche di una distribuzione statistica
- Numeri indici (Parte 6)
- Statistica bivariata:
  - Dipendenza
  - Regressione
  - Correlazione
- Probabilità
- Inferenza

PARTE 1
- Concetti introduttivi
- Rilevazione dei fenomeni statistici:
  - Rilevazione campionarie
  - Questionario e tecniche di somministrazione
  - Fonti statistiche
  - Matrice dei dati
- Statistica descrittiva e interferenza statistica

CONCETTI INTRODUTTIVI:

Definizione di statistica: è la disciplina che elabora i principi e le metodologie che presiedono:
- al processo di rilevazione dei dati
- Alla rappresentazione dei dati
- All'interpretazione dei risultati
- Alla generalizzazione delle evidenze osservate

Fasi dell'indagine statistica:
1) Formulazione del problema, eventualmente sotto forma d'ipotesi
2) Individuazione dei dati pertinenti
3) Programmazione della rivelazione dei dati
4) Analisi dei dati
5) Interpretazione dei risultati

Statistica nelle attività operative:
- Individuazione dello scopo dell'attività
- Analisi della situazione di partenza
- Esami degli strumenti da utilizzare
- Previsione dei risultati conseguibili
- Decisione finale

I quattro termini fondamentali della statistica:
1) Collettivo o Popolazione
2) Unità: caso individuale

UN COLLETTIVO È COMPOSTO DA N UNITÀ STATISTICHE

3) Carattere: aspetto rilevato sulle unità del collettivo
4) Modalità (di un carattere): uno dei diversi modi con cui il carattere si presente nelle unità del collettivo.

Es.

	Età	Sesso	Titolo di studio	Attività	Peso
Giorgio	37	M	Laurea	Occupato	70
Camilla	20	F	Licenza media	Disoccupato	53
Giordana	26	F	Laurea	Occupato	50
Nicolò	23	M	Diploma	Studente	67

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

- Giorgio, Camilla, Giordana e Nicolò sono le unità statistiche
- {Giorgio, Camilla, Giordana, Nicolò} = collettivo (popolazione)
- Età, sesso, titolo di studio, attività, peso sono i caratteri
- 37, 20, 26, 23 sono le modalità assunte dalle unità del collettivo per il carattere età

<u>Tipi di caratteri:</u>
Caratteri qualitativi (variabili qualitative, mutabili statistiche) - le modalità sono costituite da espressioni variabili
- Sconnessi: le modalità non sono ordinabili
  - Dicotomici
  - Politomici
- Ordinati: le modalità ordinabili
  - Rettilinei
  - Ciclici

Caratteri quantitativi (variabili quantitative) - le modalità sono costituite da numeri
- Discreti: le modalità sono quantità distinte
- Continui: le. Modalità possono assumere tutti i valori di un intervallo di numeri reali
- Trasferibili: un'unità può cedere dall'ammontare posseduto del carattere

<u>Misurazione dei caratteri qualitativi</u>:
- Osservare e registrare le modalità che si presentano nelle singole unità statistiche
- Le modalità possono essere predefinite a priori o desunte a posteriori

Carattere sconnesso: Scala nominale
Carattere ordinato: Scala ordinale
<u>Misurazione dei caratteri quantitativi</u>:
- Discreti: operazione di conteggio
- Continui: operazione di misurazione (con approssimazione)

Carattere discreto: scala proporzionale
Carattere continuo: scala ad intervalli, scala proporzionale

RILEVAZIONE DEI FENOMENI STATISTICI:

<u>Come avviene la rivelazione dei dati:</u>
- Indagine statistica
- Esperimento
- Osservazione sul campo

<u>Rivelazioni totali o parziali:</u>
Nell'indagine statistica, distinguiamo tra:
- Rivelazione totale (o censura): su tutto il collettivo
- Rilevazioni campionaria: su un sottoinsieme estratto dal collettivo detto campione -> campionamento

<u>Rilevazione campionarie:</u>
La numerosità (dimensione) del campione viene indicata con n (n<<N).
Frazione di campionamento: si definisce frazione campionamento o anche frazione sonda il rapporto.

$$f = \frac{n}{N}$$

Es. se N = 1000 e n = 10

$$f = \frac{10}{1000} = 0{,}01$$

<u>Tecniche di campionamento:</u>
Campionamento non casuale:
- Campionamento ragionato (o a scelta ragionata): lezione fatta da un team di esperti
- Campionamento per quota: selezione fatta su sottogruppi omogenei della popolazione
- Campionamento di convenienza

Campionamento causale (probabilistico):
Le unità sono selezionate con meccanismo casuale (aleatorio) e hanno tutte una probabilità nota e non nulla di essere selezionate.
Il campionamento probabilistico ha 2 vantaggi:
1) oggettività nella selezione delle unità
2) Possibilità di estendere i risultati all'intera popolazione

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Piani di gauscampionamento casuale:
Ogni procedura di estrazione casuale può essere assimilata al concetto intuitivo di estrazione di palline dall'urna.

Tecniche di estrazione casuale:
- Estrazione con ripetizione: la composizione dell'urna non varia
- Estrazione senza ripetizione: la composizione dell'urna varia

Piani di campionamento casuale
1) Campionamento semplice
2) Campionamento stratificato
3) Campionamento a grappoli
4) Campionamento a due stadi
5) Campionamento sistematico

1 - Campionamento casuale semplice
Alle unità della popolazione si assegna uguali probabilità di essere inserite nel campione
Distinguiamo tra:
- Campionamento casuale semplice con ripetizione
- Campionamento casuale semplice senza ripetizione

2 - Campionamento casuale stratificato
- La popolazione viene suddivisa in sotto-popolazioni (strati, unità di 1° livello) omogenee.
- Il campione è composto dalle unità (di 2° livello) estratte (con campionamento casuale semplice) da ciascuno strato.

3 - Campionamento casuale a grappolo
- La popolazione viene suddivisa in sotto-popolazioni (strati, unità di 1° livello)
- Viene estratto un campione casuale semplice di grappoli
- Il campione è composto da tutte le unità (di 2° livello) nei grappoli estratti.

4 - Campionamento casuale a due stadi
- La popolazione viene suddivisa in sotto-popolazioni (grappoli, di 1° livello)
- Viene estratto un campione casuale semplice di grappoli
- Il campione è composto dalle unità (di 2° livello) estratte (con campionamento casuale semplice) da ciascun grappolo estratto.

Unità di		**Unità di 2° livello**	
1° livello	**CENSITE**		**CAMPIONATE**
CENSITE	Censimento		Campionamento causale stratificato
CAMPIONATE	Campionamento casuale a grappolo		Campionamento casuale a 2 stadi

5 - Campionamento sistematico
- Data una popolazione la cui N le cui unità siano numerate 1 a N, da cui estrarre un campione di n unità.
- Sia dato il passo di campionamento:
- $$p \approx \frac{1}{f.} = \frac{N}{n} \text{ (p numero intero)}$$
- Sia r un numero (minore o uguale a p) scelto casualmente.

Definizione: Si denisce campione sistematico l'insieme delle unità contraddistinte con le etichette:

$$[r, r+p, r+2p, \ldots, r+(n-1)\cdot p]$$

Es. Dato un collettivo di 25 studenti, vogliamo estrarre un campione di 5 studenti con un campionamento sistematico.
La prima unità estratta casualmente è r = 2.
Il passo di campionamento è $\frac{25}{5} = 5$.
Il campione è formato dalle seguenti unità 2,7,12,17,22

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

<u>Acquisizione delle informazioni</u>:

L'intervista

Tra i metodi di acquisizione di dati uno dei più diusi è l'intervista.

- Intervista telefonica
- Intervista diretta o personale
- Intervista postale
- Intervista Web-Based

Questionario:

Domande (e relative risposte) vengono raccolte in un questionario

- Successione ordinata di domande
- Tutte le domande vanno sottoposte in modo identico ad un collettivo di unità statistiche.

Alcune regole per la redazione di un questionario

1- Brevità
2- Chiarezza delle domande
3- Evitare domande orientate
4- Limitare lo sforzo di memoria
5- Le domande riguardanti i dati socio-demografici dell'intervistato dovrebbe essere alla fine del questionario

Tipi di domande:

- Domande semplici (una sola risposta)
- Domande multiple (più risposte)
- Domande aperte
- Domande chiuse
- Domande miste
- Domande filtro

Alcuni aspetti sulla formulazione di domande chiuse

La scelta del numero di modalità di risposta può influire sui risultati.

Anche l'ordine di presentazione delle modalità di risposta influenza i risultati dell'indagine (polarizzazione delle risposte).

Nella rilevazione di opinioni e/o atteggiamenti si utilizzano delle scale (dette attitudinali).

<u>Tipi di scale attitudinali</u>

SCALA DI LIKERT:

Misura dell'atteggiamento (negativo o positivo), o di accordo rispetto ad una aermazione

Es. Le prospettive economiche per l'Italia sono incoraggianti

Per niente d'accordo	Poco d'accordo	Né d'accordo né in disaccordo	Abbastanza d'accordo	Totalmente d'accordo

SCALA DEL DIFFERENZIALE SEMANTICO:

Si sceglie il grado di accordo tra due aggettivi tra loro opposti

Es.

Roma è una città:

- Caotica ______ Ordinata
- Sporca _______ Pulita
- Chiusa _______ Accogliente

SCALA DI STAPEL:

Simile alla precedente, ma con riferimento ad un solo aggettivo

Es.

Roma è una città:

- Ordinata _____
- Pulita. _____
- Accogliente _____

<u>L'ISTAT</u>

- L'ISTAT (Istituto Nazionale di Statistica) è l'Ente preposto all'acquisizione dei dati tramite indagini statistiche.
- E' al vertice del Sistema Statistico Nazionale (Sistan)
- Effettua rilevazioni censuarie e campionarie

Es. di rivelazioni campionarie

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

- Indagine sulle forze lavoro (effettuata ogni 3 mesi)
- Indagine sui consumi delle famiglie

Censimenti permanenti:

- A partire dal 2018, i censimenti vengono sostituiti dai censimenti permanenti.

https://www.istat.it/it/censimenti-permanenti

- I censimenti permanenti avranno cadenza annuale, biennale o triennale e riguarderanno un campione rappresentativo della popolazione di interesse.
- Sono estesi a tutte le aree tematiche
  - popolazione e abitazioni
  - imprese
  - istituzioni non profit
  - istituzioni pubbliche
  - agricoltura
- Il Censimento permanente della popolazione e delle abitazioni (partito il 1° ottobre 2018) riguarda un campione di un milione e quattrocentomila famiglie circa.
- I dati campionari verranno integrati con le fonti amministrative

Altre fonti statistiche

- Eurostat (Ufficio Statistico dell'Unione Europea)
- Banca d'Italia
- BCE (Banca Centrale Europea)
- Censis (Centro Studi Investimenti Sociali)

Matrice dei dati:

- La matrice dei dati raccoglie le informazioni rilevate sulle unità di un collettivo (o di un campione) su un insieme di caratteri.
- Ogni riga si riferisce ad una unità
- Ogni colonna si riferisce ad un carattere

Cod. ID	Sesso	Stato civile	Provincia	Titolo di studio	Professione	Numero comp. famiglia	Reddito Lordo Annuo
1	F	Nubile	RM	Laurea	Impiegato	1	35000
2	F	Nubile	FR	Diploma	Disoccupato	3	10000
3	M	Coniugato	LT	Laurea	Impiegato	4	40000
4	F	Coniugato	FR	Laurea	Libero professionista	3	90000
5	M	Celibe	RI	Laurea	Libero professionista	1	90000
6	F	Coniugato	RM	Diploma	Casalinga	4	15000
7	M	Celibe	RM	Licenza media	Pensionato	5	30000
8	F	Nubile	FR	Laurea	Disoccupato	3	20000
9	F	Coniugato	FR	Laurea	Libero professionista	3	50000
10	M	Celibe	RM	Laurea	Pensionato	2	70000

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

STATISTICA DESCRITTIVA E INFERENZA STATISTICA:
Statistica descrittiva: Procedure descrittive per individuare le configurazioni essenziali dei dati.
Inferenza statistica: Procedure che permettono la generalizzazione dei risultati osservati sul campione casuale.

PARTE 2
Distribuzione statistiche e rappresentazione i grafiche
Operatore sommatoria:
Siano a1,a2,...,ai,...,aN un insieme di numeri.
Definizione:
La sommatoria di ai per i che va da 1 a N è così definita

$$\sum_{i=1.}^{N} ai = a1 + a2 + \ldots + a1 + aN$$

Es. a1 =4, a2 =3, a3 =5 allora $\sum_{i=1}^{3} ai = 4 + 3 + 5 = 12$

Proprietà:

- $\sum_{I=1}^{N} (ai + bi) = \sum_{i=1.}^{N} ai + \sum_{i=1}^{N} bi$
- $\sum_{i=1}^{N.} c * ai = \sum_{i=1}^{N} ai$
- $\sum_{i=1}^{N} c = c * N$

Dove c è una costante (non varia al variare di i)
Esercizio:
Dati

N = 4,
c = 2,
a1 = 4, a2 = 9, a3 = 6, a4 = 3,
b1 = 5, b2 = 7, b3 = 3, b4 = 4

dimostrare le proprietà 1-3.

Per comodità, disponiamo i numeri in forma tabellare:

I	**ai**	**bi**	**ai + bi**	**c**	**c * ai**
1	4	5	9	2	8
2	9	7	16	2	18
3	6	3	9	2	12
4	3	4	7	2	6
Tot.	**22**	**19**	**41**	**8**	**44**

Ad esempio:

$$\sum_{i=1}^{4.} ai + \sum_{i=1}^{4.} b1 = 22 + 19 = 41 = \sum_{i=1}^{4} (ai + bi)$$

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Altri esercizi:
Con gli stessi valori dell'esempio precedente, dimostrare che:

$$\sum_{i=1}^{N} (ai * bi) = / \sum_{i=1.}^{N} ai * \sum_{i=1}^{N} bi$$

$$\sum_{i=1}^{N} a^2 i = /(\sum_{i=1}^{N} ai)^2$$

Regole per l'arrotondamento:
Arrotondare un numero signica ridurre il numero di cifre decimali (quelle dopo la virgola)

Valore originaria	1 decimale	2 decimali
12, 422	12,4	12,42
11,237	11,2	11,24
11,262	11,3	11,26
10,251	10,3	10,25
10,257	10,3	10,26
10,255	10,3	10,26

Contenuto PARTE 2
- Distribuzioni statistiche disaggregate
- Distribuzioni di frequenze:
  - Distribuzioni di frequenze per modalità singole
  - Distribuzioni di frequenze per classi di valori
  - Distribuzioni di frequenze cumulate
  - Densità di frequenza
- Altre forme di distribuzioni
- Rappresentazioni grafiche:
  - Diagramma a torta
  - Diagramma a barre
  - Diagramma cartesiano Istogramma
  - Cartogramma
  - Funzione di ripartizione

DISTRIBUZIONI STATISTICHE DISAGGREGATE:
Definizione: La distribuzione (statistica) semplice disaggregata di un carattere è l'elenco delle modalità osservate, unità per unità, nel collettivo in esame.

Unità	X
u1	x1
u2	x2
//	//
ui	xi
//	//
uN	xN

N.B. E' detta anche:
- Distribuzione (statistica) semplice untaria
- Serie di osservazionI

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Es: Distribuzione unitaria semplice di un collettivo in base all'età

Unità	**Età**
1	45
2	26
3	37
4	37
5	27
6	64
7	37
8	38
9	49

Es. Distribuzione unitaria semplice di un collettivo in base al sesso

Unità	**Sesso**
1	M
2	M
3	M
4	F
5	M
6	F
7	M
8	F
9	F

Distribuzione statistica multipla disaggregata:
Definizione - Pariamo di distribuzione (statistica) multipla consideriamo più caratteri.

Unità	**X**	**Y**	**Z**
u1	x1	y1	z1
u2	x2	y2	z2
//	//	//	//
ui	xi	yi	Z3
//	//	//	//
uN	xN	yN	zN

N.B. E' detta anche distribuzione (statistica) unitaria multipla

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Es: Distribuzione unitaria multipla per i caratteri Età, Sesso e Statura su una popolazione

Unità	Età	Sesso	Statura
1	45	M	181
2	26	M	174
3	37	M	171
4	37	F	180
5	27	M	182
6	64	F	184
7	37	M	166
8	38	F	172
9	49	F	177

NB. Ogni colonna corrisponde alla distribuzione unitaria di un carattere
DISTRIBUZIONE DI FREQUENZE:
Dalla distribuzione unitaria alla distribuzione di frequenze -
Consideriamo la seguente serie di osservazioni relativa al numero di gli di un collettivo di 22 famiglie:

2, 2, 1, 4, 3, 1, 1, 3, 1, 3, 1, 2, 2, 2, 3, 0, 0, 1, 2, 1, 2, 1

Possiamo rappresentare la distribuzione in maniera più sintetica con la seguente distribuzione di frequenze (per modalità singole):

Figli	Famiglia
0	2
1	8
2	7
3	4
4	1
	22

Frequenza di una modalità:
Distinguiamo tra:
- frequenza assoluta: numero di volte che una determinata modalità si verica nel collettivo di riferimento
- frequenza relativa: frequenza assoluta rapportata al totale delle unità statistiche della popolazione
- frequenza percentuale: frequenza relativa moltiplicata per 100

Distribuzione di frequenze:
Definizione - La distribuzione di frequenze è una tabella in cui ad ogni modalità di una variabile fa corrispondere la rispettiva frequenza
La distribuzione di frequenze si dice:
- semplice (un carattere)
- multipla (più caratteri)

A seconda del tipo di frequenza avremo:
- distribuzione di frequenze assolute
- distribuzione di frequenze relative
- distribuzione di frequenze percentuali

Distribuzione di frequenze assolute:

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Dato il carattere X con k modalità (x1,...,xi,...,xk) osservato su n unità statistiche.
Siano n1, . . . , ni , . . . , nk le frequenze assolute delle k modalità.
Distribuzione di frequenze assolu

X	ni
xi	n1
//	//
xi	ni
//	//
xK	nk
	N

Dove:

$$n1 + \ldots + ni + \ldots + nk = \sum_{i=1}^{k} ni = N$$

Esempio: Distribuzione degli immatricolati della Facoltà di Scienze Politiche, Sociologia, Comunicazione per corso di laurea - a.a. 2016-17

Distribuzione di frequenze relative:
Indichiamo con

f1 = n1/N, ..., fi = ni/N, ..., fk = nk/N

le frequenze relative (o proporzioni) delle k modalità.
Distribuzione di frequenze relative

X	f1
x1	$f_1 = n_1/N$
//	//
xi	$f_i = n_i/N$
//	//
xk	$f_k = n_k/N$
	1

dove:

$$f1 + \ldots + fi + \ldots + fk = \sum_{i=1}^{k} fi = 1 \qquad 0 \leq fi \leq 1$$

Distribuzioni di frequenze percentuali:
Indichiamo con

p1 =f1·100,...,pi =fi ·100,...,pk =fk ·100l

e frequenze percentuali delle k modalità.
Distribuzione di frequenze percentuali

X	pi
x1	$p_1 = f_1 \cdot 100$
//	
xi	$p_i = f_i \cdot 100$

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

X	**pi**
//	
xk	$pk = fk \cdot 100$
	100

Esempio: Distribuzione di un collettivo di famiglie per numero di figli

Figli	**Famiglia**	**fi**	**pi**	
	0	2	0,09	9
	1	8	0,36	36
	2	7	0,32	32
	3	4	0,18	18
	4	1	0,05	5
		22	1,00	100

Ad esempio, la percentuale di famiglie con 2 gli è pari al 32%
Esempio: Confronto in termini assoluti e relativi
Residenti per titolo di studio in due aree:

A

Titolo di studio.	ni
Licenza elementare.	759
Licenza media.	1113
Diploma.	2459
Laurea.	1669
	6000

B

Titolo di studio.	ni
Licenza elementare	95
Licenza media.	163
Diploma.	181
Laurea	361
	800

In termini assoluti ci sono molti più laureati in A che in B.

Residenti per titolo di studio in due aree:

A

Titolo di studio.	ni	fi
Licenza elementare.	759.	0,13
Licenza media.	1113.	0,19
Diploma.	2459.	0,41
Laurea.	1669	0, 28
	6000.	1,01

B

Titolo di studio.	ni.	fi
Licenza elementare	95.	0,12
Licenza media.	163.	0,20
Diploma.	181.	0,23
Laurea	361.	0,45
	800.	1,00

In termini relativi è esattamente il contrario.
Dalla distribuzione unitaria alla distribuzione di frequenze per classi di modalità:
Consideriamo un collettivo di 48 aziende sulle quali è stato osservato il numero di addetti:
14, 20, 13, 3, 18, 9, 10, 20, 16, 12, 15, 2, 2, 5, 17, 10, 1, 8, 2, 9, 3, 20, 14, 15, 9, 10, 19, 7, 19, 7, 10, 16, 9, 8, 9, 8, 6, 9, 20, 7, 19, 20, 16, 20, 3, 4, 1, 8
Possiamo rappresentare la distribuzione in maniera più sintetica: distribuzione di frequenze per classi di modalità.

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Addetti	Aziende
5 |- 10	15
10 |- 15	8
15 |—| 20	16
	48

Suddivisione in classi del carattere
Quando il carattere quantitativo discreto presenta un elevato numero di modalità distinte, oppure è continuo, può essere utile suddividere in classi le modalità del carattere.

Indicazioni per la determinazione delle classi
1- numero di classi abbastanza piccolo da fornire un'adeguata sintesi
2- numero di classi abbastanza grande da mantenere l'informazione
3- classi tra loro disgiunte
4- le classi devono comprendere tutte le possibili modalità del carattere
5- le classi dovrebbero avere la stessa ampiezza (se possibile)

Tipologie di classi
Siano:
- ci l'estremo superiore della classe
- ci–1 l'estremo inferiore della classe (ci–1 < ci)

Casi possibili:
- classe chiusa a sinistra e aperta a destra: ci–1 ⊢ ci
- classe aperta a sinistra e chiusa a destra: ci–1 ⊣ ci
- classe chiusa a sinistra e a destra: ci–1 ⊢⊣ ci
- classe aperta a sinistra e a destra: ci–1 – ci

Tipologie di classi. (Notazione alternativa)
Siano:
- ci l'estremo superiore della classe
- ci–1 l'estremo inferiore della classe (ci–1 < ci)

Casi possibili:
- classe chiusa a sinistra e aperta a destra: [ci–1, ci)
- classe aperta a sinistra e chiusa a destra: (ci–1, ci]
- classe chiusa a sinistra e a destra: [ci–1, ci]
- classe aperta a sinistra e a destra: (ci–1, ci)

Distribuzione di frequenze per classi di valori:
La distribuzione di frequenze per classi di valori associa a ciascuna classe la rispettiva frequenza (assoluta).

Classe	ni
$c_0 \vdash c_1$	n1
//	//
$c_{i-1} \vdash c_i$	ni
//	//
$c_{k-1} \vdash\dashv c_k$	nk
	N

E' immediato derivare le distribuzioni di frequenze relative e percentuali per classi di valori.

Esempio: Distribuzione di un collettivo di famiglie per classi di consumo familiare mensile

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Consumo	Famiglie	fi	pi
1⊢2	117	0,39	39
2⊢3	77	0,26	26
3⊢⊣6	39	0,13	13
	300	1,00	100

Le famiglie con un consumo superiore o uguale a 1000 e inferiore a 3000euro sono il 65% del totale.

Frequenze assolute cumulate:
Definizione - La frequenza assoluta cumulata di una modalità del carattere misura il numero di casi che presentano un valore non superiore a quella modalità:

$$N_i = n_1 + n_2 + \ldots + n_i, \; i = 1, \ldots, k$$

Distribuzione di frequenze assolute cumulate.

X	ni	
x1	n1	$N_1=n_1$
x2	n2	$N_2=n_1+n_2$
//	//	
xi	ni	$N_i = n_1+\ldots+n_i$
//	//	
xk	nk	$N_k = N$

Frequenze relative cumulate

$$F_i = f_1 + f_2 + \ldots + f_i, \; i = 1, \ldots, k$$

Distribuzione di frequenze relative cumulate.

X	p1 percentuali	Pi cumulate
x1	p1	$P_1 = p_1$
x2	p2	$P_2 = p_1 + p_2$
//	//	//
xi	pi	$P_i = p_1+\ldots+p_i$
//	//	//
xk	pk	$P_k = 100$

Esempio: Distribuzione di un collettivo di famiglie per numero di figli

Figli	Famiglie	fi	pi	Ni	Fi	Pi
0	2	0,09	2		0,09	9

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Figli	Famiglie	fi	pi	Ni	Fi	Pi
1	8	0,36	10		0,45	45
2	7	0,32	17		0,77	77
3	4	0,18	21		0,95	95
4	1	0,05	22		1,00	100
	22	1,00	100			

Le famiglie che hanno fino a 2 figli sono 17 (77 % del totale).
Esempio: Distribuzione di un collettivo di famiglie per classi di consumo mensile

Consumo	Famiglie	fi	pi	Ni	Fi	Pi
0⊢1	67	0,22	22	67	0,22	22
1⊢2	117	0,39	39	184	0,61	61
2⊢3	77	0,26	26	261	0,87	87
3⊢⊣6	39	0,13	13	300	1,00	100
	300	1,00	100			

Le famiglie il cui consumo è di almeno 1000euro al mese sono il 78% del totale.
Ampiezza e densità di frequenza di una classe:
Sia data una generica classe di estremi ci−1 e ci.
Ampiezza della classe: di =ci −ci−1
Densità di frequenza:

$$hi = \frac{ni}{di}$$

[ni = frequenza assoluta nella classe
di = ampiezza della classe]
La densità di frequenza sono sempre confrontabili tra le classi.

Esempio: Distribuzione degli addetti di un'azienda per classi di età

Classi	ni	di	hi
18 ⊢ 25	25	7	3,57
25 ⊢ 30	20	5	4,00
30 ⊢ 40	21	10	2,10
40 ⊢ 50	21	10	2,10
50 ⊢⊣ 64	33	14	2,36
	120		

N.B.
La classe con la frequenza più alta non coincide necessariamente con quella. On la densità di frequenza più alta.

ALTRE FORME DI DISTRIBUZIONE:
Distribuzione di quantità:

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Definizione - Schema con cui si associa a ogni modalità di un carattere trasferibile X il totale dello stesso posseduto dalle unità che presentano quella data modalità di X.
Osservazione: Il carattere per il quale si calcola la distribuzione di quantità può essere diverso rispetto a quello su cui si fa la classicazione.
Esempio: Distribuzione di un collettivo di famiglie per numero di figli.

N° figli	Distribuzione di frequenza (n° di famiglie)	Distribuzione di quantità (n° totale di figli)
1	10	10
2	12	24
3	8	24
4	2	8
	32	66

Esempio: Distribuzione del reddito mensile (in migliaia di euro) di un collettivo di famiglie classicate per numero di figli.

N° figli	Distribuzione di frequenza (n° di famiglie)	Distribuzione di quantità (reddito totale)
1	10	27
2	12	30
3	8	16
4	2	5
	32	78

Serie storiche
Definizione - Si parla di serie storiche (temporali) quando si misura un fenomeno in determinati istanti di tempo.
Esempio:
Popolazione residente in Italia (in migliaia)

Anno	Popolazione
2012	59394
2013	59685
2014	60783
2015	60796
2016	60666
2017	60589

Serie territoriali

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Definizione - Si parla di serie territoriali (geografiche o spaziali) quando le modalità rappresentano entità territoriali ossia rappresentano ad esempio nazioni, regioni, ripartizioni geografiche, città, etc.
Esempio: Popolazione residente in Italia nel 2017, per area geografica (in migliaia)

Area	Popolazione
Nord - ovest	16104
Nord - est	11637
Centro	12068
Mezzogiorno	20781
Isole	6710

RAPPRESENTAZIONE GRAFICHE:

- La rappresentazione grafica di una distribuzione statistica può essere più informativa della distribuzione stessa
- Il tipo di grafico deve essere scelto in base:
  - al tipo di distribuzione
  - al tipo di carattere

Il diagramma a torta (pie chart):
Utilizzo principale
Distribuzione di frequenze:
- caratteri qualitativi sconnessi
- serie territoriale

Le "fette della torta" sono calcolate in base alla seguente formula:

$$a_i^\circ = 360° \times \frac{n_i}{N} = 360° \times f_i$$

- N: numerosità della popolazione
- ni: frequenza della i-ma modalità
- fi: frequenza relativa della i-ma modalità
- a○: angolo della fetta della i-ma modalità

[Esempio: Popolazione residente in Italia nel 2017, per area geografica (in migliaia)]

Diagramma a barre:
Utilizzo principale:
Distribuzione di frequenze:
- caratteri qualitativi (sconnessi o ordinati)
- caratteri quantitativi discreti
- serie storica
- serie territoriale

L'altezza di ciascuna barra è proporzionale alla frequenza della modalità corrispondente.
[Esempio 1 - 2: Spettatori per tipo di evento nel 2017 (in migliaia)]
[Esempio: Distribuzione degli immatricolati della Facoltà di Scienze Politiche, Sociologia, Comunicazione per corso di laurea - a.a. 2016-17]
[Esempio: Famiglie per numero di figli]

Diagramma cartesiano:
Utilizzo principale:
Distribuzione di frequenza:
- caratteri quantitativi discreti
- serie storica

Osservazione: I punti sul piano possono essere uniti da una spezzata, per evidenziare l'andamento del fenomeno.

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

[Esempio: Appartamenti per numero di stanze]
[Esempio: Popolazione residente in Italia (in migliaia)]
[Esempio: Percentuale di canzoni di successo di durata inferiore a 2 minuti e mezzo]

Istogramma:
Impiego: Distribuzioni di frequenze per classi di valori di un carattere quantitativo
L'istogramma è composto da rettangoli adiacenti, la cui area è data dalla seguente formula:

$$\text{base} \times \text{altezza} = \text{ampiezza della classe} \times \text{densità di frequenza} = d_i \cdot h_i = d_i \cdot \frac{n_i}{d_i} = n_i = \text{Area}$$

[Esempio: Temperature medie / Classi della stessa ampiezza]
[Esempio: Temperature / Classi di ampiezza differente]

Cartogramma:
I cartogrammi sono particolari rappresentazioni grafiche per serie territoriali.
[Esempio: Parità di potere di acquisto in Europa (2018)]

Funzione di ripartizione per caratteri quantitativi discreti
Definizione:
Dato un carattere quantitativo discreto X con k modalità, osservato su un collettivo di N unità, la funzione di ripartizione F(x) è così definita:

$$F(x) = \begin{cases} 0, & \text{per } x < x_1 \\ F_i, & \text{per } x_i \le x < x_{i+1} \quad (i=1,\dots,k-1) \\ 1, & \text{per } x \ge x_k \end{cases}$$

dove F_i è la frequenza relativa cumulata per la modalità i-ma
[Esempio: Famiglie per numero di figli]

Funzione di ripartizione per caratteri quantitativi continui:
Definizione: Dato un carattere quantitativo continuo X con modalità raggruppate in k classi, osservato su un collettivo di N unità, la funzione di ripartizione F(x) è così definita:

$$F(x) = \begin{cases} 0, & \text{per } x < c_0 \\ F_{i-1} + \frac{f_i}{d_i}(c_i - 1) & \text{per } c \le x < c \quad (i=1,\dots,k) \\ 1, & \text{per per } x \ge c_k \end{cases}$$

dove F_{i-1} è la frequenza relativa cumulata per la classe i – 1.
[Esempio: Quotazioni di titoli azionari]

PARTE 3
- Medie analitiche
  - Media aritmetica
  - Proprietà della media aritmetica
  - Media quadratica
  - Media armonica
  - Media geometrica
- Media aritmetica per distribuzione di frequenze
- Medie lasche
  - La mediana
  - Proprietà della mediana
  - La mediana per distribuzioni di frequenze
  - Quartili
  - I quartili per distribuzioni di frequenze
  - Valore centrale
  - Moda

Motivazioni:
Sintesi della distribuzione di un carattere attraverso una modalità ”tipica” con l'obiettivo di:

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

- valutare l'ordine di grandezza di un fenomeno;
- effettuare comparazioni tra le manifestazioni di uno stesso fenomeno in tempi, luoghi o situazioni diverse;
- comunicare tale informazione.

Tale modalità tipica è detta media

Le medie si distinguono in:
1) medie analitiche
2) medie lasche

1 - Medie analitiche
Medie che richiedono opportune operazioni matematiche in cui si tiene conto di tutti i valori delle modalità.
Si applicano solo a caratteri di tipo quantitativo.
- media aritmetica
- media armonica
- media geometrica
- media quadratica

2 - Medie lasche
Medie che utilizzano alcuni valori specici della distribuzione, senza coinvolgere nel calcolo tutte le modalità del carattere.
Possono essere applicate anche a caratteri qualitativi ordinati.
- mediana
- quantili
- valore centrale
- moda (anche per caratteri sconnessi)

MEDIE ANALITICHE:
- Data una distribuzione statistica disaggregata, o serie di osservazioni, di un carattere quantitativo osservata su un collettivo di dimensione N:

x1, x2, ...,N

- La media aritmetica è la somma dei termini diviso per N e la indichiamo con μ:

$$\mu = \frac{x1 + x2 + \ldots + xn}{N} = \frac{1}{N} \sum xi$$

Esempio: Numero di libri letti
Su un collettivo di 6 studenti abbiamo rilevato il numero di libri letti nell'ultimo mese:

4,3,5,4,2,3

La media aritmetica è:

$$\mu = \frac{1}{6}(4 + 3 + 5 + 4 + 2 + 3) = \frac{21}{6} = 3{,}5$$

Numero complessivo di libri letti dagli studenti nel collettivo, diviso il numero di studenti.
Domanda: Come si interpreta questo valore medio?
Proprietà della media aritmetica:
Siano x(1) e x(N) il minimo ed il massimo valore della distribuzione:
1) Condizione di internalità - $x(1) \leq \mu \leq x(N)$

Esempio:
Riprendendo l'esempio precedente (numero di libri letti per studente nell'ultimo mese):

$x(1) = 2 < 3{,}5 < 5 = x(6)$

Domande
- In quale situazione $x(1) = \mu = x(N)$?
- Può capitare che $x(1) \leq \mu = x(N)$
- E invece che $x(1) = \mu \leq x(N)$?

N

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

2) Equiripartizione - $\sum_{i=1} xi = N\mu$

La media aritmetica è la quantità che, se sostituita a ciascuna modalità osservata nel carattere, ne lascia invariata la somma.
Esempio:
Riprendendo l'esempio precedente: 4 + 3 + 5 + 4 + 2 + 3 = 21 = 6 · 3,5

3) La somma algebrica degli scarti dalla media aritmetica è nulla: $\sum_{i=1}^{n} (xi - \mu) = 0$

Esempio:
Riprendendo l'esempio precedente

Studente	Libri	xi - μ
1	4	0,5
2	3	-0,5
3	5	1,5
4	4	0,5
5	2	-1,5
6	3	-0,5
	21	0,0

4) La somma algebrica degli scarti dalla media aritmetica al quadrato è minima.

$$\sum_{i=1}^{n} (xi - \mu)^2 = \min_{c} \sum_{i=1}^{n} (xi - c)^2$$

Esempio:

Studente	Libri	$xi - \mu$	$(xi - \mu)^2$
1	4	0,5	0,25
2	3	-0,5	0,25
3	5	1,5	2,25
4	4	0,5	0,25
5	2	-1,5	2,25
6	3 -0.5		0,25
	21	0,0	5,50

Per esercizio: sostituire a μ = 3,5 un valore qualsiasi (ad esempio, c=4) e confrontare il risultato con quello della tabella.

5) Linearità della media

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Se si trasformano i termini x1,x2,...,xN secondo la funzione:

yi = a + bxi   i=1,2,...,N (a e b costanti qualsiasi)

allora:

μY =a+bμX

Esempio:
- 5 turisti italiani fanno un prelievo da un bancomat all'aeroporto di Fiumicino, prima di partire per New York.
- Le somme prelevate sono le seguenti:

100, 250, 50, 80, 100

Una volta arrivati a destinazione, cambiano gli euro in dollari.
1 - Qual è la media dei prelievi in euro?
2 - Tenuto conto della commissione di 2 dollari per l'operazione e del fatto che il cambio è di 1,2$ per euro, qual è la media in dollari?
- La media dei prelievi in euro è
μe = 116
- Per il calcolo della media in dollari, abbiamo due alternative:
  (I) Potremmo trasformare tutti i valori secondo la funzione:

y =-2+1,2·x

  ottenendo

118, 298, 58, 94, 118

  e poi calcolare la media di questi valori (μ$ = 137,2)
  (II) In maniera più diretta, applichiamo alla media la formula vista prima:

μ$ = -2+1,2·μe = 137,2

Per esercizio, controllare i calcoli.
6) Associatività della media:
Dato un collettivo statistico di N unità suddiviso in L sottoinsiemi disgiunti di numerosità

N(1),N(2),…,N(L) $\left(\sum_{l=1}^{L} N^{(l)} = N\right)$

E media aritmetiche:

μ1,μ2,…,μL

la media aritmetica del collettivo può essere calcolata con la seguente formula:

$$\mu = \frac{1}{N}(\mu 1 \cdot N(1) + \mu 2 \cdot N(2) + ... + \mu L \cdot N(L))$$

Esempio:

Sesso	**Reddito**
F	1500
F	1800
M	1500
M	2400
M	3200
M	1600
Tot.	12000

- La media del reddito è di 2000 euro.
- Il reddito medio delle 2 donne è 1650 euro, quello dei 4 uomini è 2175.
La media del reddito può essere calcolata in questo modo:

$$\mu = \frac{N^F \mu^F + N^M \mu^M}{N} = \frac{21650 + 42175}{6} = 2000$$

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

Esempio: Immatricolati alla Facoltà di Scienze Politiche, Sociologia, Comunicazione - a.a. 2016-17

Corso di Laurea	Immatricolati
Comunicazione Pubblica e d'Impresa	272
Comunicazione, Tecnologie e Culture Digitali	330
Cooperazione internazionale e Sviluppo	109
Relazioni Economiche Internazionali	79
Scienze dell'Amministrazione e dell'Organizzazione	94
Scienze e Tecniche del Servizio Sociale	165
Scienze Politiche e Relazioni Internazionali	563
Sociologia	511

$$\mu = \frac{2123}{8} = 265{,}375$$

La media quadratica:
Data una serie di osservazioni di un carattere quantitativo osservata su un collettivo di dimensione N: x1, x2, …,N
La media quadratica è la radice quadrata della media aritmetica dei quadrati dei termini della distribuzione:

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari