

METODOLOGIA STATISTICA E CLASSIFICAZIONE DEI DATI

1.1 La Statistica

La Statistica è la scienza che raccoglie, elabora ed interpreta i dati (informazioni) relativi ad un dato fenomeno oggetto di osservazione. L'oggetto di osservazione ovvero l'oggetto dell'indagine statistica può riguardare tutti i fenomeni collettivi e cioè quei fenomeni che sono costituiti da una pluralità di elementi di cui è necessario analizzare e sintetizzare particolari aspetti.

La Statistica si divide in tre sezioni principali: la Statistica metodologica, la Statistica descrittiva e la Statistica inferenziale.

La *Statistica metodologica*, si occupa di studiare i metodi logici, matematici e tecnici per raccogliere e classificare i fenomeni di massa che interessano.

La *Statistica descrittiva*, si occupa di predisporre i metodi per estrarre ed esporre in forma sintetica, le informazioni rilevanti da una grande mole di dati.

La *Statistica inferenziale*, ha il compito di analizzare un campione estratto da una certa popolazione di cui non sia possibile avere una conoscenza esaustiva, e da esso trarre indicazioni sulle caratteristiche dell'intera popolazione che non potrebbe essere analizzata singolarmente. L'estensione delle indicazioni rilevate sul campione possono essere estese all'intera popolazione utilizzando la teoria della probabilità.

1.2 Nomenclatura statistica

Sia la Statistica metodologica che la Statistica descrittiva e la Statistica inferenziale condividono le seguenti definizioni di base di alcuni termini usati in questo lavoro:

Definizioni:	Significato:
Universo statistico	L'Universo statistico è l'insieme costituito da tutti gli elementi interessati all'indagine statistica (Unità statistiche). L'Universo statistico è indicato con le lettere S, T, U ecc. L'Universo statistico ha come sinonimi: <ul style="list-style-type: none">- Popolazione;- Popolazione statistica;- Collettivo statistico;
Unità statistica	Le Unità statistiche sono gli elementi costituenti l'Universo statistico. Esse sono entità reali o virtuali rispetto alle quali si vuole compiere l'indagine statistica. Le Unità statistiche sono indicate con le lettere A, B, C ecc. Le Unità statistiche possiedono, più caratteristiche o proprietà (Caratteri statistici) non necessariamente tutte interessate al fenomeno oggetto di osservazione. L'identificazione di una Unità statistica con le sue caratteristiche oggetto di osservazione, costituisce il <i>Caso statistico</i> .
Carattere statistico	Il Carattere statistico è una caratteristica delle Unità statistiche osservabile e/o misurabile. Notiamo ancora una volta che una stessa Unità statistica, può avere più Caratteri statistici diversi.

	<p>I Caratteri statistici sono indicati con le lettere X, Y, Z ecc.</p> <p>Il Carattere statistico ha come sinonimi:</p> <ul style="list-style-type: none"> - Caratteristica; - Fenomeno; - Aspetto; - Proprietà. <p>I Caratteri statistici possono essere classificati sulla base dei modi con cui possono essere rilevati e/o misurati (Modalità).</p> <p>Una classificazione dei Caratteri statistici che tiene conto delle <i>modalità</i> di rilevazione e/o misurazione li suddivide in due grandi categorie:</p> <ul style="list-style-type: none"> - <i>Caratteri qualitativi</i> se le modalità con cui vengono rilevati individuano attributi o indicano delle qualità. In questo caso i Caratteri sono chiamati <i>Mutabili statistiche</i>; - <i>Caratteri quantitativi</i> se le modalità con cui vengono rilevati sono enumerazioni o misure e quindi sono esprimibili mediante numeri interi o indicano delle quantità. In questo caso i Caratteri sono chiamati <i>Variabili statistiche</i>.
Modalità	<p>Le Modalità sono i vari modi in cui può venire rilevato/misurato un carattere statistico.</p> <p>Le Modalità di rilevazione e/o di misura producono informazioni sui Caratteri rilevati o misurati (Dati Statistici)</p>
Dato statistico	<p>Il Dato statistico è una informazione su un Carattere di una Unità statistica oggetto di rilevazione e/o misurazione. Esso rappresenta il risultati delle <i>Misure o Rilevazioni</i> effettuate sui Caratteri delle Unità statistiche attraverso una Modalità prestabilita.</p> <p>Se il Carattere da rilevare è qualitativo le relative rilevazioni produrranno Dati detti <i>Dati categorici</i>.</p> <p>I Dati categorici possono essere così classificati:</p> <ul style="list-style-type: none"> - Dati categorici binomiali - Dati categorici nominali; - Dati categorici ordinali. <p>Se il Carattere da misurare è quantitativo allora le relative misure produrranno Dati detti <i>Dati numerici</i>.</p> <p>I Dati numerici possono essere così classificati:</p> <ul style="list-style-type: none"> - Dati numerici continui; - Dati numerici discreti. <p>I Dati numerici sono espressi mediante numeri</p>
Intensità	<p>L'intensità indica il risultato della misura effettuata su un determinato Carattere quantitativo secondo una Modalità prestabilita.</p>

Frequenza	La Frequenza indica il numero di volte in cui una data Modalità si è presentata all'interno del fenomeno oggetto di studio.
Indagini totali	Le indagini totali, sono indagini statistiche riferite ad Universi statistici le cui unità statistiche sono reali e finite rispetto alle quali l'indagine può essere estesa a tutte le Unità statistiche.
Indagini parziali	Le indagini parziali sono quelle indagini effettuate solo su una parte della popolazione detta <i>campione</i> .

Esempio

Individuiamo come *Universo statistico S* l'Insieme degli studenti iscritti all'Università della Calabria nell'A.A. 2006/07 .

Le *Unità statistiche* di *S* sono quindi tutte le persone che nel periodo 1° Settembre 2006 – 31 Luglio 2007 sono stati iscritti all'Università della Calabria. L'insieme delle Unità statistiche può essere rappresentato con la seguente tabella.

Tabella delle Unità statistiche oggetto di rilevazione

Cognome e nome degli studenti

Rossi Francesco
 Bianchi Emanuela
 Carli Giuseppe
 Guidi Antonio
 De Marchi Anna
 Franchi Roberta

.....

Per ogni studente (Unità statistica) è possibile definire più *Caratteri*:

- Sesso;
- Comune di residenza;
- Età;
- Altezza;
- Titolo di Studio;
- Titoli di Studi del padre;
- Soluzione di una verifica;
- Voto di Maturità;
- Ecc.

Il *Sesso* è un Carattere rilevabile mediante due *Modalità*: Maschio (M) o Femmina (F). Tali modalità sono attributi qualitativi non ordinabili dello studente, pertanto il Sesso è una *Mutabile statistica* i cui Dati sono riportati nella seguente tabella.

Tabella dei Dati di rilevazione del Sesso

<i>Cognome e nome degli studenti</i>	<i>Sesso</i>
Rossi Francesco	M
Bianchi Emanuela	F
Carli Giuseppe	M
Guidi Antonio	M
De Marchi Anna	F
Franchi Roberta	F

.....

Tali Dati sono classificabili come Dati categorici binomiali

Il *Comune di residenza* è un Carattere che può assumere come *Modalità*: Cosenza, Reggio Calabria, Rende, ecc. Tali modalità sono attributi qualitativi non ordinabili dello studente che possono assumere svariati valori. Pertanto il Comune di Residenza è una *Mutabile statistica* i cui Dati sono riportati nella seguente tabella.

Tabella dei Dati di rilevazione del Comune di residenza

<i>Cognome e nome degli studenti</i>	<i>Comune di residenza</i>
Rossi Francesco	Cosenza
Bianchi Emanuela	Reggio Calabria
Carli Giuseppe	Rende
Guidi Antonio	Reggio Calabria
De Marchi Anna	Rende
Franchi Roberta	Cosenza

Tali Dati sono classificabili come Dati categorici nominali.

Il *Titolo di Studio* è un Carattere che può assumere come *Modalità*: XXXXXXXXXXXXXXXX
Ecc. Tali modalità sono attributi qualitativi dello studente e quindi il Titolo di Studio è una *Mutabile statistica*.

Le modalità con cui si può rilevare il Titolo di Studio, possono essere ordinate sulla base degli anni necessari al conseguimento dello stesso:

Ordinamento del Titolo di Studio per anni necessari al conseguimento

<i>Titolo di Studio</i>	<i>Anni necessari al conseguimento</i>
XXXXXXXX	0
XXXXXX	1
	2
	3
	4
	5

Pertanto il Titolo di Studio è una Mutabile statistica ordinabile.

Tabella dei Dati di rilevazione dei titoli di studio del padre

<i>Cognome e nome degli studenti</i>	<i>Titolo di studio del padre classificato secondo la tabella precedente</i>
Rossi Francesco	0
Bianchi Emanuela	2
Carli Giuseppe	4
Guidi Antonio	3
De Marchi Anna	3
Franchi Roberta	1

Tali Dati sono classificabili come Dati categorici ordinali.

La *Soluzione di una verifica* è un Carattere quantitativo degli studenti dimessi che può essere rilevato con varie *Modalità*: giorni, ore, ecc. Tali modalità sono attributi quantitativi dello studente e quindi la soluzione di una verifica è una *Variabile statistica* i cui dati statistici sono riportati nella seguente tabella:

Tabella dei Dati di rilevazione della soluzione di una verifica

<i>Cognome e nome degli studenti</i>	<i>Soluzione in giorni</i>	<i>Soluzione in ore</i>
Rossi Francesco		
Bianchi Emanuela		
Carli Giuseppe		
Guidi Antonio		
De Marchi Anna		
Franchi Roberta		
.....		

Tali Dati sono classificabili come Dati numerici continui.

Il voto di maturità è un Carattere che può essere rilevato con più *Modalità*: un qualsiasi numero intero compreso tra 36 e 60 per i voti espressi in sessantesimi ed un qualsiasi numero intero compreso tra 60 e 100 per i voti espressi in centesimi ad esempio;

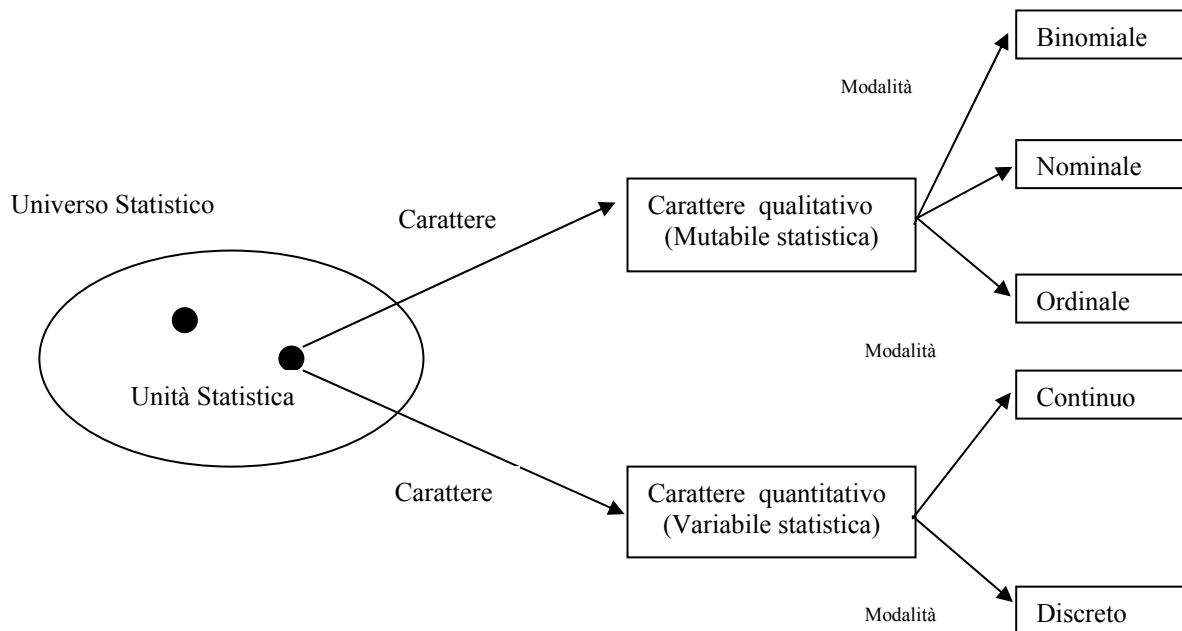
Tali modalità sono attributi quantitativi dello studente e quindi il Voto di maturità è una *Variabile statistica* i cui dati statistici sono riportati nella seguente tabella:

Tabella dei Dati di rilevazione dei Voti di maturità degli studenti

<i>Cognome e nome degli studenti</i>	<i>Voto di Maturità in sessantesimi</i>	<i>Voto di maturità in centesimi</i>
Rossi Francesco	60	100
Bianchi Emanuela	48	80
Carli Giuseppe	48	80
Guidi Antonio	51	85
De Marchi Anna	60	100
Franchi Roberta	57	95
.....		

Tali Dati sono classificabili come Dati numerici discreti.

Grafico di rappresentazione sintetica della nomenclatura statistica



1.3 Tipologia e classificazione dei Dati

Il Dato rappresenta il valore che può assumere una determinata Modalità di un Carattere. Più precisamente il Dato è il risultato di una misura effettuata su una Unità statistica per rilevarne un determinato Carattere attraverso una Modalità prestabilita.

I Dati possono essere classificati come *Dati grezzi* o *Dati elementari* quanto sono rilevati in modo oggettivo direttamente sulle Unità statistiche senza ulteriori elaborazioni. Sono classificati come *Dati elaborati* quanto sono costruiti a partire dai Dati grezzi mediante processi di elaborazione in modo da renderlo significativo ai fini della acquisizione delle informazioni necessarie nei processi decisionali.

1.4 Tipologie di Dati grezzi

I Dati grezzi o elementari possono venire classificati nelle seguenti *Tipologie*:

- Dati categorici:
 - binomiali;
 - nominali;
 - ordinali.
- Dati numerici:
 - discreti;
 - continui.

1.4.1 Dati categorici binomiali:

I Dati categorici binomiali sono dati riferiti a Caratteri qualitativi la cui modalità può assumere solo due valori.

Esempi:

Riferiti al carattere "Sesso" la modalità è: Maschio/Femmina;

Riferiti al carattere "Evento" la modalità è: Vero/Falso;

Riferiti al carattere "Scommessa" la modalità è: Vinta/Persa.

1.4.2 Dati categorici nominali:

I Dati categorici nominali sono dati riferiti a Caratteri qualitativi le cui modalità possono assumere più valori per i quali non è possibile definire in modo naturale una relazione d'ordine.

Esempi:

Riferiti al carattere "Professione" la modalità può assumere i valori: Operaio, Impiegato, Dirigente, Medico, Avvocato, Ingegnere, Commercialista Ecc.;

Riferiti al carattere "Nazionalità" la modalità può assumere i valori: Francese, Svizzera, Ecc.;

Riferiti al carattere "Stato Civile" la modalità può assumere i valori: Celibe, Nubile, Coniugato;

1.4.3 Dati categorici ordinali:

Sono dati riferiti a Caratteri qualitativi le cui modalità possono assumere più valori per i quali è possibile definire in modo naturale una relazione d'ordine.

Esempi:

Riferiti al carattere "Titolo di studio" i valori della modalità possono essere ordinate sulla base degli anni di studio necessari a conseguire il titolo: Nessuno, Licenza elementare, Diploma di scuola media inferiore, Diploma di scuola media superiore, Laurea di primo livello, Laurea specialistica Ecc.;

1.4.4 Dati numerici discreti:

Sono dati riferiti a Caratteri quantitativi le cui modalità generano solo numeri interi e non frazionabili. I Dati numerici discreti sono di norma generati dalla modalità di *enumerazione*.

Esempi:

Riferiti al carattere "N. figli" i valori della modalità sono: 0, 1, 2, 3, 4 Ecc.;

Riferiti al carattere "N. Esami superati" i valori della modalità sono: 0, 1, 2, 3, Ecc.;

Riferiti al carattere "N. abitanti" i valori della modalità sono: 1, 2, ..1.236,... 3.425,.... 22.345 Ecc.;

1.4.5 Dati numerici continui:

Sono dati riferiti a Caratteri quantitativi la cui modalità può assumere tutti i valori numerici compresi in un dato intervallo detto Dominio della Variabile statistica. I Dati numerici continui sono generati dalla modalità di *misurazione*.

Esempi:

Riferiti al carattere "altezza di un individuo" i valori della modalità sono tutti numeri compresi nell'intervallo dei valori da 1,00 a 2,50 metri;

Riferiti al carattere "temperatura corporea" i valori della modalità sono tutti numeri compresi nell'intervallo dei valori da 35 a 42 gradi centigradi;

1.4.5.1 Scale di intervalli e di rapporti

Le scale a intervalli consentono di formare delle graduatorie; si caratterizzano però per il fatto che l'intervallo tra due posizioni successive resta costante per tutta l'estensione della scala. Se la scala consente di identificare anche una posizione corrispondente alla mancanza di misura, cioè a zero, si ha una scala di rapporti. In questo caso va osservato che i numeri impiegati per indicare le posizioni sulla scala possiedono sia la proprietà ordinale (oltre alla simmetrica ed alla

transitiva presenti già nei precedenti tipi di scale), sia quella cardinale: è perciò possibile effettuare con esse le operazioni aritmetiche.

Esempi:

E' una scala a intervalli quella del termometro, in cui l'origine (0°) è arbitraria, mentre sono scale di rapporti quelle del sistema metrico decimale (misure di lunghezza, volume ecc.).