

# Parte 1

Introduzione:  
Statistica e metodologia della ricerca

# Contenuti

- Introduzione alla ricerca statistica
- Problemi reali e problemi statistici
- Fasi di una indagine statistica
- Elementi introduttivi del software Minitab

N.B. Le slide di **approfondimento** non sono materia di esame

# Cos'è la Statistica?

- La Statistica è una disciplina scientifica che studia metodi e strumenti per acquisire conoscenza su un'ampia varietà di problemi e fenomeni in molti ambiti applicativi (scienze sociali, naturali, economiche; ingegneria, ....).
- I fenomeni di interesse della Statistica sono detti **collettivi**. Un fenomeno è detto collettivo quando possiamo studiarlo solo attraverso l'osservazione e l'analisi di un **insieme** di sue manifestazioni.
- La Statistica non affronta, analizza o risolve *casi individuali, singoli*, ma cerca di pervenire a **conclusioni generali** su **fenomeni collettivi** e alla discussione critica della loro validità a partire dalla conoscenza aggregata risultante da **un insieme di casi singoli (popolazione o campione)** accomunati da **regole e definizioni**, sui quali è osservato e analizzato un insieme di **caratteristiche (variabili)**.

## *Esempio:*

### **Controllo statistico di un processo produttivo**

Nelle aziende di produzione assume un ruolo rilevante il controllo statistico di processo finalizzato a controllare l'insieme di caratteristiche che identificano il livello di produzione. Risponde a domande del tipo: i parametri della produzione rispettano standard nazionali e internazionali previsti da normative di riferimento? Qual'è il grado di soddisfazione dell'utente finale? ...

⇒ Necessità di osservare periodicamente **campioni** di prodotti.

⇒ Necessità di raccogliere informazioni su un insieme di caratteristiche (**variabili**) diverse: caratteristiche di prodotto, costi di produzione, soddisfazione dell'utente, ... .

## Esempi:

- Un rubinetto idraulico di un sistema di irrigazione dovrebbe secondo progetto erogare  $1 \text{ m}^3$  di acqua al minuto. Si vuole studiare la variabilità nell'erogazione dell'acqua.
  - È di interesse lo studio dell'effetto del degrado di malte e mattoni sui materiali di costruzione.
  - Si vuole verificare la tensione di rottura di barre d'acciaio di differenti diametri via prova di trazione.
  - Si vuole studiare la variabilità di metodiche applicate da diverse aziende produttrici di calcestruzzo.
- ⇒ Necessità di osservare **popolazioni** o **campioni** di prodotti, provini, aziende, ...
- ⇒ Necessità di raccogliere informazioni su una serie di caratteristiche diverse (**variabili**): peso dell'acqua erogata, stato tensionale, tipi di calcestruzzo, ...

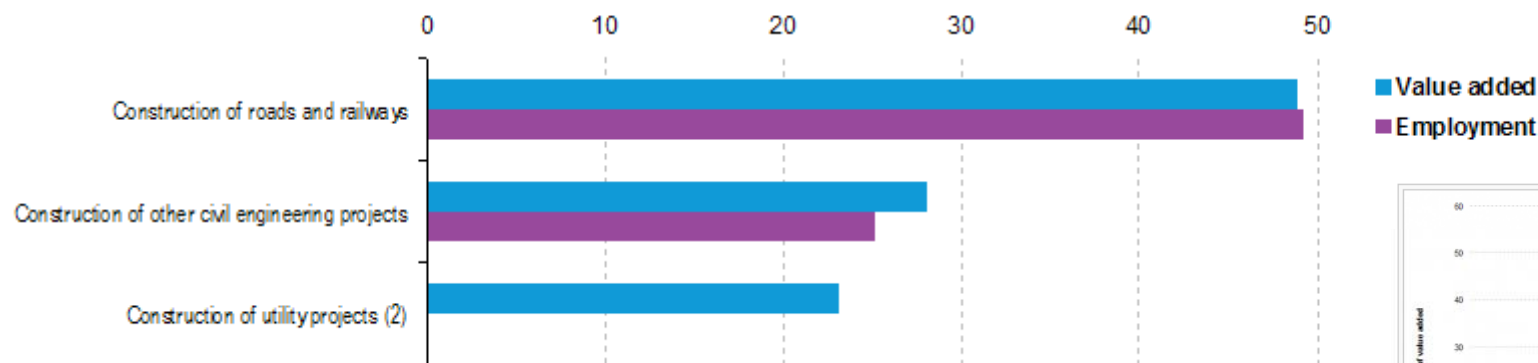
# Cos'è la Statistica?

- L'osservazione (rilevazione, misurazione, ...) e l'analisi statistica di variabili in una popolazione o campione, genera una serie di **dati**. I **dati statistici** non sono semplicemente numeri, ma numeri che forniscono **informazioni** su un problema o su una certa realtà in studio.
- La Statistica fornisce ragionamenti e metodi per **produrre**, **analizzare** e **interpretare** dati statistici.
- La Statistica:
  - trasforma i dati in informazioni utili per **descrivere**, **monitorare** uno specifico argomento di interesse;
  - fornisce metodi per **ricercare relazioni** tra fenomeni;
  - interviene nelle situazioni nelle quali occorre **assumere decisioni** in **condizioni di incertezza**.

## Esempio: Descrizione e monitoraggio

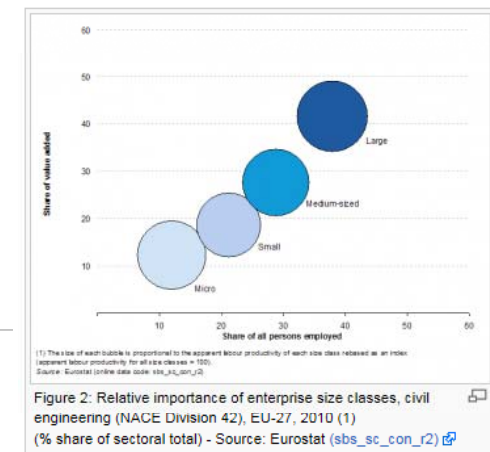
L'Unione Europea si avvale di numerosi **indicatori statistici** al fine di monitorare e di predisporre politiche di sviluppo nel settore dell'ingegneria civile. Sul sito dell'Eurostat (**Ufficio Statistico dell'Unione Europea**) sono pubblicati periodicamente dati, analisi e rapporti di monitoraggio che forniscono un quadro obiettivo sullo stato e sull'evoluzione del settore. L'analisi e l'interpretazione di tali indicatori (presentati con tabelle, grafici o altre visualizzazioni), consentono di valutare se l'Europa si sta muovendo nella giusta direzione e con sufficiente rapidità secondo gli obiettivi definiti dalle politiche europee.

### Sectoral analysis of civil engineering (NACE Division 42), EU-27, 2010 (1) (% share of sectoral total).



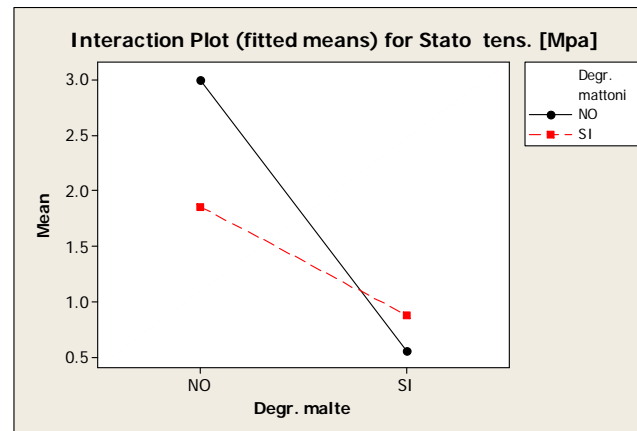
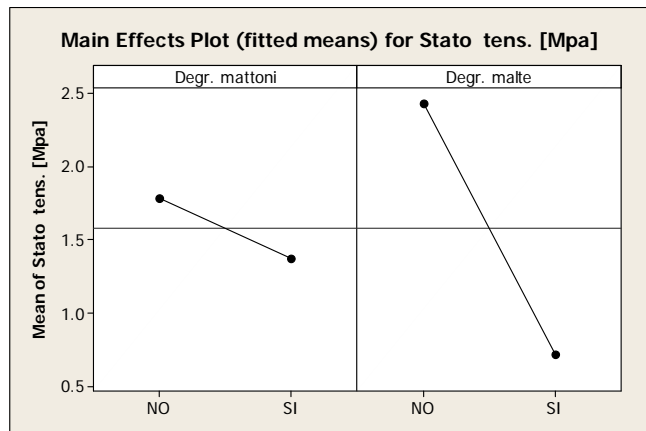
### Sectoral analysis of key indicators, civil engineering (NACE Division 42), EU-27, 2010 A.

	Number of enterprises	Number of persons employed	Turnover	Value added	Personnel costs
	(thousands)	(thousands)	(EUR million)	(EUR million)	(EUR million)
<b>Civil engineering</b>	101.5	1 630.0	355 000	72 200	51 900
Construction of roads and railways	35.0	802.9	134 000	35 300	26 000
Construction of utility projects	25.6	·	52 500	16 700	12 800
Construction of other civil engineering projects	40.6	410.0	99 000	20 300	13 100



## Esempio: Ricerca di relazioni tra fenomeni

In uno studio sui materiali di costruzione viene misurato lo stato tensionale di 16 carote per capire se esiste un effetto causato dal degrado dei mattoni e delle malte. L'obiettivo è di stabilire se il degrado di mattoni e di malte, così come la loro interazione, causa una variazione significativa nello stato tensionale.



Source	DF	Seq SS	Adj SS	Adj MS	F	P
Degr. mattoni	1	0.6765	0.6765	0.6765	56.13	<b>0.000</b>
Degr. malte	1	11.8164	11.8164	11.8164	980.45	<b>0.000</b>
Degr. mattoni*Degr. malte	1	2.1683	2.1683	2.1683	179.91	<b>0.000</b>
Error	12	0.1446	0.1446	0.0121		
Total	15	14.8058				
S = 0.109782    R-Sq = 99.02%    R-Sq(adj) = 98.78%						



## **Esempio: Prendere decisioni in condizioni di incertezza**

Un'azienda produttrice di barre d'acciaio per opere edilizie, avendo necessità di **compiere investimenti in nuovi impianti** che consentano una maggiore capacità produttiva, vuole compiere **previsioni sulle vendite** nei tre anni successivi.

Le **condizioni di incertezza** possono presentarsi ad esempio per:

- la limitatezza delle osservazioni (indagini campionarie);
- la complessità del fenomeno o della realtà da indagare che si traduce ad esempio nella difficoltà di individuare tutte le variabili di interesse;
- imprecisione nelle misurazioni o nelle rilevazioni delle variabili;
- instabilità del fenomeno nel tempo, dovuta ad esempio dalla dipendenza da altri fenomeni (es. crisi economica, variazioni dei prezzi delle materie prime, ...).

# Problemi reali e problemi statistici

- Nella ricerca scientifica vale il “**Principio della parsimonia scientifica**”, che implica l'accettazione di un certo grado d'incertezza nella ricerca e la scelta di modelli di rappresentazione della realtà in studio, il più possibile semplici.
- Nello studio di un problema di interesse è quindi necessaria una semplificazione del problema nelle sue **linee essenziali** per poter procedere nell'approfondimento secondo metodi statistici opportuni.
- La semplificazione consiste nel tradurre il problema di interesse in un **problema statistico** definendo una struttura formale (**modello**) e una **griglia concettuale** (definizioni, classificazioni, ...).

# Problemi reali e problemi statistici

- Per un dato problema di interesse, la ricerca statistica intraprende quindi una **indagine** specificando il **modello** (struttura formale) e la **griglia concettuale** (definizioni, classificazioni, ...) all'interno dei quali saranno generati e avranno validità i **dati statistici** informativi sulla realtà in studio.
- Il modello e la griglia concettuale risultano definiti in base agli **obiettivi** della ricerca, la **tipologia dello studio** (trasversale, longitudinale; osservazionale, sperimentale; ...), la **natura della rilevazione** (censuaria, campionaria; diretta, basata su fonti già disponibili; ...), la **tecnica dell'indagine** (questionario, diario, ...), i **vincoli temporali e le risorse**.
- I dati statistici non sono quindi semplicemente numeri, ma numeri che forniscono **informazioni** su un problema o realtà di interesse all'interno di uno **specifico contesto di ricerca** e che **necessitano di essere interpretati all'interno di quel contesto**.

## Esempio:



I risultati di qualsiasi indagine statistica vanno interpretati tenendo conto del **contesto** entro il quale sono stati prodotti.

**Definizione della popolazione, della tipologia di studio, della tecnica di indagine, ...**

**L'Indagine sulla fiducia delle imprese di costruzione**, prevista nell'ambito del programma armonizzato relativo alle inchieste congiunturali presso le imprese e i consumatori approvato dalla Commissione Europea, è finalizzata alla rilevazione di opinioni, quali giudizi e aspettative a breve termine (3 mesi), sui principali aspetti economici delle imprese di costruzione (p.es. giudizi sul portafoglio ordini e sull'attività di costruzione, previsioni sugli ordini etc.). L'unità di rilevazione e di analisi è l'impresa. L'universo di riferimento è rappresentato dalle imprese attive con almeno tre addetti appartenenti al settore delle costruzioni così come individuato nella sezione F dell'ateco 2007. L'indagine è mensile ed è effettuata presso un panel longitudinale di imprese con almeno 3 addetti; da novembre 2011 è condotta con tecnica CATI. Il disegno campionario utilizzato è di tipo casuale stratificato.

## Esempio:

PROSPETTO 3. CLIMA DI FIDUCIA DELLE IMPRESE DI COSTRUZIONE TOTALE E PER SETTORE  
Aprile-agosto 2014, indici destagionalizzati (base 2005=100) e saldi destagionalizzati

	2014				
	Apr	Mag	Giu	Lug	Ago
<b>COSTRUZIONI</b>					
Clima di fiducia	75,1	73,6	81,3	83,0	77,0
Giudizi su ordini e/o piani di costruzione	-50	-54	-44	-45	-48
Attese sull'occupazione	-24	-23	-22	-14	-22
<b>COSTRUZIONE DI EDIFICI</b>					
Clima di fiducia	57,8	55,2	64,6	64,0	62,6
Giudizi su ordini e/o piani di costruzione	-63	-64	-56	-59	-57
Attese occupazione	-31	-35	-25	-24	-29
<b>INGEGNERIA CIVILE</b>					
Clima di fiducia	92,9	85,0	83,4	87,5	83,0
Giudizi su ordini e/o piani di costruzione	-10	-23	-27	-21	-26
Attese occupazione	-7	-9	-8	-7	-11
<b>LAVORI DI COSTRUZIONE SPECIALIZZATI</b>					
Clima di fiducia	87,8	93,0	95,4	107,4	
Giudizi su ordini e/o piani di costruzione	-48	-47	-36	-32	
Attese occupazione	-20	-13	-21	-7	

Definizioni di variabili,  
classificazioni, scale di misura, ...

## Glossario

**Clima di fiducia delle imprese per settore:** è elaborato tramite media aritmetica semplice dei saldi destagionalizzati delle domande ritenute maggiormente idonee per valutare l'ottimismo/pessimismo delle imprese. Il risultato è poi riportato a indice in base 2005.

**Settori delle costruzioni:** i settori delle costruzioni oggetto d'indagine, con i relativi codici della classificazione Ateco 2007, sono la costruzione di edifici (41), l'ingegneria civile (42) e i lavori di costruzione specializzati (43). La *costruzione di edifici* include lavori generali per la costruzione di edifici di qualsiasi tipo. L'*ingegneria civile* comprende i lavori generali per la costruzione di opere di ingegneria civile quali autostrade, strade, ponti, gallerie, ferrovie, campi di aviazione, porti ed altre opere idrauliche, nonché la costruzione di sistemi di irrigazione e di fognatura, condotte e linee elettriche, impianti sportivi all'aperto, eccetera. I *lavori di costruzione specializzati* comprendono attività specializzate (quali l'infissione di pali, i lavori di fondazione eccetera), attività di finitura e completamento degli edifici e attività di installazione di tutti i tipi di servizi, necessarie al funzionamento della costruzione.

# Le fasi di una indagine statistica

Per l'esecuzione di una indagine statistica è importante procedere ad una attenta **pianificazione**.

Le **fasi** di una indagine statistica possono essere sintetizzate nel modo seguente:

- Definizione degli obiettivi della ricerca
- Definizione della tipologia di studio
- Definizione della popolazione di interesse
- Definizione delle variabili di interesse e delle scale di misura
- Definizione della fonte dei dati
- Definizione della tecnica di indagine (strumento e modalità di raccolta dati)
- Selezione del campione (per indagini campionarie)
- Rilevazione dei dati e organizzazione dei dati raccolti
- Analisi statistica e presentazione dei risultati
- Interpretazione e utilizzazione dei risultati della ricerca

# Le fasi di una indagine statistica

## ▪ Definizione degli obiettivi della ricerca

Nello studio di un problema di interesse, è necessario specificare e dichiarare gli **obiettivi**, al fine di definire correttamente la natura delle informazioni da raccogliere e gli strumenti statistici con i quali esaminare i dati. Gli obiettivi devono essere quindi dettagliati in modo da circoscrivere l'ambito dell'indagine, evitare equivoci definitivi, individuando con esattezza il territorio e il periodo dell'indagine. Devono inoltre essere fattibili ed agevoli da investigare.

Definire gli obiettivi significa: **delimitare precisamente cosa interessa da cosa non interessa** ricordando che più ampio è l'arco degli argomenti trattati, maggiori divengono le complessità da affrontare sul piano concettuale statistico e operativo; definire se interessa descrivere un fenomeno nella sua **componente statica o in quella dinamica**; specificare se interessa **confrontare** i risultati con informazioni relative ad altre realtà territoriali o nel tempo; specificare eventuali **ipotesi** da sottoporre a verifica.

**Conoscenze preliminari** e **analisi della letteratura** sull'argomento di interesse sono fondamentali per una corretta definizione degli obiettivi dell'indagine.

# Le fasi di una indagine statistica

## ▪ Definizione della tipologia di studio

La tipologia dello studio può differenziarsi in particolare con riferimento:

- al collettivo da esaminare;
- alle modalità temporali di rilevazione;
- alla periodicità dello studio;
- alle modalità di intervento del ricercatore;
- alle finalità della ricerca.



Con riferimento al  
collettivo da esaminare



**Indagini globali (censuarie):**

viene osservata tutta la popolazione di interesse

**Indagini parziali (campionarie):** viene osservata una parte o campione della popolazione di interesse

Con riferimento alle  
modalità temporali di  
rilevazione



**Studi trasversali:** viene effettuata una unica rilevazione in un istante temporale

**Studi longitudinali (prospettici, retrospettivi, misure ripetute):** per ogni unità di rilevazione vengono rilevati dati in corrispondenza di più istanti temporali

Con riferimento alla  
periodicità dello studio



**Studi occasionali:**

lo studio è condotto una sola volta nel tempo

**Studi ripetuti:**

lo studio è ripetuto periodicamente nel tempo

Con riferimento alle  
modalità di intervento  
del ricercatore



**Studi osservazionali:** c'è l'osservazione passiva della realtà di interesse senza interventi del ricercatore

**Studi sperimentali:** il ricercatore interviene modificando la realtà in studio mediante un esperimento

Con riferimento alle  
finalità della ricerca



**Studi descrittivi o esplorativi:** l'interesse del ricercatore si limita alla descrizione della realtà senza formulare ipotesi di ricerca

**Studi esplicativi o analitici:** l'interesse del ricercatore è verificare ipotesi di ricerca (presenza di relazioni tra fenomeni, ...)

# Le fasi di una indagine statistica

- **Definizione della popolazione di interesse**

Si definisce **unità statistica** l'elemento (individuo, struttura, ente, azienda, prodotto, provino di un materiale, ...) sul quale viene effettuata la rilevazione o la misurazione di una o più caratteristiche rilevanti per il problema in studio.

Si definisce **popolazione** l'insieme delle unità statistiche interessate dal problema in studio.

# Le fasi di una indagine statistica

- Definizione delle variabili di interesse e delle scale di misura

Si definisce **variabile (o carattere)** una caratteristica rilevata o misurata sulle unità statistiche. Le variabili assumono **valori differenti (modalità)** nelle varie unità statistiche e possono essere:

**Variabili quantitative**

le modalità sono numeri reali

**Variabili qualitative o categoriali**

le modalità sono attributi non numerici

Le **variabili qualitative** possono essere:

**Variabili non ordinabili o sconnesse**  
le modalità non sono ordinabili

**Esempio:**

Variabile	Modalità
Presenza di scorie nella miscela di calcestruzzo	- si - no
Strato della sovrastuttura stradale di conglomerati bituminosi	- di usura - di collegamento o binder - di base

**Variabili ordinabili**  
le modalità sono ordinabili

**Esempio:**

Variabile	Modalità
Reazione al fuoco di materiali di costruzione	- materiali incombustibili - mat. combustibili non o difficilmente infiammabili - mat. combustibili normalmente infiammabili - mat. combustibili facilmente infiammabili
Livello di gocciolamento durante la combustione di materiali di costruzione	- nessun gocciolamento - lento gocciolamento - elevato gocciolamento

Le **variabili quantitative** possono essere:

### **Variabili discrete**

possono assumere solo valori appartenenti all'insieme dei numeri naturali

**Esempio:** numero di dipendenti di una azienda, numero di anomalie di un elemento strutturale, ...

### **Variabili continue**

possono assumere qualsiasi valore contenuto in un dato intervallo reale

**Esempio:** peso, temperatura, voltaggio, ...

# Le fasi di una indagine statistica

## ▪ Definizione delle variabili di interesse e delle scale di misura

Si definisce **scala di misura** il tipo di misurazione o l'insieme delle modalità adottato per l'osservazione di una variabile.

Alla scelta della scala di misura contribuiscono:

- gli obiettivi dell'indagine;
- la tipologia della variabile.

L'individuazione della scala di misura e della tipologia delle variabili sono fondamentali per una corretta selezione delle procedure di analisi statistica da applicare.

La scala di misura utilizzata per rilevare una variabile può essere modificata successivamente in fase di analisi dei dati. In ogni caso occorre utilizzare, quando possibile e coerentemente con gli scopi della ricerca, le **classificazioni standard disponibili** (es. Istat, Eurostat, ...), procedendo eventualmente a disaggregazioni o ad aggregazioni di codici in modo da potersi comunque riportare ad una classificazione nota, assicurando così la confrontabilità dei risultati conseguiti con quelli desumibili da altre fonti.

## Scala nominale

- Si individua un insieme di **categorie o attributi** (modalità) **non ordinabili** nelle quali viene rilevata o classificata la variabile d'interesse.
- Le variabili misurate utilizzando la scala nominale sono dette anche **nominali**.

### Esempio

Statistical Classification of Economic Activities in the European Community, NACE Rev. 2 (2008)

### Classificazione delle attività economiche Ateco 2007

n.e.c. : not elsewhere classified			
Division	Group	Class	
<b>SECTION F — CONSTRUCTION</b>			
41	41.1		Construction of buildings
			Development of building projects
	41.2	41.10	Development of building projects
		41.20	Construction of residential and non-residential buildings
42	42.1		Civil engineering
			Construction of roads and railways
		42.11	Construction of roads and motorways
		42.12	Construction of railways and underground railways
	42.2	42.13	Construction of bridges and tunnels
			Construction of utility projects
		42.21	Construction of utility projects for fluids
	42.9	42.22	Construction of utility projects for electricity and telecommunications
			Construction of other civil engineering projects
		42.91	Construction of water projects
43	43.1	42.99	Construction of other civil engineering projects n.e.c.
			Specialised construction activities
			Demolition and site preparation
		43.11	Demolition
		43.12	Site preparation
	43.13	Test drilling and boring	

F	COSTRUZIONI	
42	INGEGNERIA CIVILE	
42.11.00	Costruzione di strade, autostrade e piste aeroportuali	
42.12.00	Costruzione di linee ferroviarie e metropolitane	
42.13.00	Costruzione di ponti e gallerie	
42.21.00	Costruzione di opere di pubblica utilità per il trasporto di fluidi	
42.22.00	Costruzione di opere di pubblica utilità per l'energia elettrica e le telecomunicazioni	
42.91.00	Costruzione di opere idrauliche	
42.99.01	Lottizzazione dei terreni connessa con l'urbanizzazione	
42.99.09	Altre attività di costruzione di altre opere di ingegneria civile nca	

- Un caso particolare di variabili nominali è quello delle variabili **dicotomiche** o **binarie** così chiamate perché assumono solo **due** modalità:

***Esempio:***

Variabile	Modalità
Condizione occupazionale	- occupato - non occupato
Tipologia edificio	- residenziale - non residenziale
Presenza di anomalie in un elemento strutturale	- si - no

- Le variabili nominali che assumono più di due modalità possono essere chiamate anche **polinomiali**.



## Scala ordinale

- Si individua un insieme di **categorie o attributi** (modalità) **ordinabili in ordine crescente o decrescente** nelle quali viene rilevata o classificata la variabile d'interesse.
- Le variabili misurate utilizzando la scala ordinale sono dette anche **ordinali**.

### *Esempio*

Il **grado di urbanizzazione di un territorio (variabile)** può essere rilevato secondo la seguente **classificazione** (Eurostat) che prevede categorie con percentuali crescenti di popolazione che vive in raggruppamenti urbani e nei centri urbani:

- Zona scarsamente popolata (zona rurale)
- Area di densità intermedia (città e periferie / piccola area urbana)
- Zona densamente popolata (città / grande area urbana)

## Scala numerica

- Si individua un insieme di **valori numerici** (modalità) per misurare la variabile d'interesse.
- Le scale numeriche possono essere **ad intervallo** (**variabili intervallari**) se lo **zero** della scala delle misure è convenzionale e non indica l'assenza del fenomeno misurato.

*Esempio:* scale numeriche per misurare la latitudine, la longitudine, la temperatura

- Le scale numeriche possono essere **di rapporto** (**variabili proporzionali**) se lo **zero** della scala delle misure è univocamente definito.

*Esempio:* scale numeriche per misurare l'età, la durata, il peso, ...

# Le fasi di una indagine statistica

## ▪ Definizione delle fonti dei dati

Le informazioni sulle variabili di interesse per il problema in studio possono derivare da:

- **Fonti primarie:** le informazioni vengono raccolte ex novo mediante una particolare tecnica di indagine (intervista o osservazione diretta, ...)
- **Fonti secondarie:** le informazioni derivano da studi già effettuati in precedenza, come ad esempio dati delle fonti statistiche ufficiali (ISTAT, Sistema Statistico Nazionale, Eurostat, ...), studi pubblicati su articoli scientifici, dati disponibili via Web, ... .

# Le fasi di una indagine statistica

- **Definizione della tecnica di indagine (costruzione dello strumento e delle modalità di raccolta dati)**

Quando la **fonte dei dati** è **primaria** è necessario **definire** la **tecnica di indagine**. Quando la fonte dei dati è **secondaria** è necessario **conoscere** la **tecnica di indagine**.

Con il termine **tecnica di indagine** si intende **la procedura di contatto delle unità statistiche** interessate dalla rilevazione **e di reperimento delle informazioni** oggetto di interesse.

La scelta della tecnica di indagine più idonea a raccogliere le informazioni oggetto della ricerca è uno degli aspetti di maggiore importanza nella pianificazione e nell'esecuzione di una indagine ed è strettamente connessa a considerazioni quali: il problema indagato, gli obiettivi dell'indagine, l'eventuale strategia di campionamento, le risorse dell'indagine (in termini finanziari e di personale), i vincoli temporali dell'indagine, le implicazioni della tecnica di indagine prescelta sulla qualità dei dati, in termini di mancate risposte e di errori di misura.

Le principali **tecniche di indagine** in uso per condurre una rilevazione, sono le seguenti:

- **Intervista diretta (o faccia a faccia)**: le informazioni sulle unità statistiche sono raccolte mediante intervista diretta effettuata da un rilevatore che compila un **questionario** in cui sono riportate le risposte.
- **Intervista telefonica**: le informazioni sulle unità statistiche sono raccolte mediante intervista telefonica effettuata da un rilevatore che compila un **questionario** in cui sono riportate le risposte.
- **Questionario postale autocompilato**: le informazioni sulle unità statistiche sono raccolte mediante autocompilazione di un **questionario** recapitato a mezzo posta.
- **Diario**: E' un particolare tipo di **questionario** strutturato appositamente per registrare eventi frequenti quali spese o attività quotidiane. L'organizzazione di tale strumento è tale da permettere la registrazione degli eventi nel momento in cui avvengono in modo tale da non dover ricorrere ad uno sforzo di memoria nello svolgimento di una intervista di tipo classico.
- **Dati amministrativi**: le informazioni sulle unità statistiche provengono da **archivi** di dati amministrativi, ovvero informazioni raccolte e conservate da istituzioni pubbliche o private per scopi amministrativi e non statistici. Ad esempio sono dati amministrativi quelli raccolti su persone ed imprese ai fini fiscali, pensionistici, anagrafici o giuridici.

*Continua*

- **Osservazione diretta:** le informazioni sulle unità statistiche sono raccolte dal rilevatore per mezzo dei propri sensi o mediante strumenti di misurazione (applicazioni in geologia, telerilevamento,...).
- **Tecniche miste:** le informazioni sulle unità statistiche sono raccolte mediante la combinazione di due o più tecniche di indagine.  
Esempi di tecniche miste:  
Indagine postale + indagine diretta sui non rispondenti all'indagine postale  
Indagine telefonica + indagine diretta su coloro che non possiedono il telefono  
Diario + intervista diretta  
Prima intervista diretta e successive con modalità telefonica  
Dati amministrativi + controllo campionario con questionario postale autocompilato
- **Nuove tecnologie:** le informazioni sulle unità statistiche sono raccolte mediante intervista diretta o telefonica con compilazione di un **questionario** contenuto nel computer, cosicché le risposte sono registrate direttamente su supporto magnetico, o mediante autocompilazione di un **questionario** via Web.  
Esempi di nuove tecnologie:  
CATI (Computer Assisted Telephone Interviewing)  
CAPI (Computer Assisted Personal Interviewing)  
CAWI (Computer Assisted Web Interviewing)

Molte tecniche di indagine utilizzano il **questionario** come strumento di raccolta dei dati.

Si distinguono diverse tipologie di questionario:

- Questionario **cartaceo**
- Questionario **informatizzato** (Es. indagini CAPI, CATI, CAWI)

Si distinguono diverse tipologie di **compilazione** del questionario:

- Autocompilazione o autosomministrazione
- Compilazione mediante intervista diretta o telefonica
- Somministrazione mista mediante intervista e autocompilazione

Il questionario deve essere uno **strumento standardizzato**, ovvero domande e risposte devono essere identiche per tutte le unità statistiche di rilevazione affinché le informazioni raccolte siano confrontabili fra loro.

Per la realizzazione di un questionario è necessario:

- definire esattamente quali sono i **temi** che interessano l'indagine escludendo quelli che non sono di interesse primario;
- preparare la **lista delle variabili** (e non direttamente le domande) da raccogliere rispetto ai temi di interesse identificati in precedenza;
- preparare un piano provvisorio delle **analisi statistiche** da compiere per accertarsi di raccogliere tutte le informazioni necessarie.

Le **fasi** di redazione di un questionario sono le seguenti:

- Stabilire la successione logica dei temi trattati (le **sezioni** del questionario)
- Definire la **sequenza di domande** per ciascuna sezione
- Formulare i **quesiti**
- Organizzare le **risposte**
- **Verificare** il questionario (pre-test, indagini pilota)



# Le fasi di una indagine statistica

- **Selezione del campione (per indagini campionarie)**

Quando l'indagine è campionaria è necessario definire la **strategia di campionamento** e procedere quindi alla selezione del campione.

Il **campione** rappresenta un insieme di unità statistiche tratte dalla popolazione di interesse.

La definizione delle modalità di estrazione del campione e della sua dimensione prende il nome di **strategia di campionamento** ed è basata sulla ben consolidata **teoria dei campioni**.

Selezionare solo un campione implica ovviamente avere informazioni parziali rispetto alla possibilità di osservare l'intera popolazione. Tuttavia, se la selezione del campione viene effettuata in modo corretto (**campionamento casuale o probabilistico**), è possibile **estendere, generalizzare** i risultati dal campione alla popolazione e misurare il livello di precisione delle informazioni campionarie (**inferenza statistica**).

# Le fasi di una indagine statistica

## ▪ Rilevazione dei dati e organizzazione dei dati raccolti

Sono **fasi operative** in cui si procede a raccogliere i dati in base alle scelte effettuate relativamente alla tipologia dello studio, alla fonte dei dati, alla tecnica di indagine, e a organizzare i dati in un formato (banca dati, database, data set) utile per procedere all'analisi statistica.

Prima dell'analisi statistica dei dati, sulle informazioni raccolte possono essere effettuate procedure di:

- **revisione**, mediante tecniche manuali o automatiche al fine di evidenziare la presenza di mancate risposte (**dati mancanti o missing**) o di errori, rispetto ai quali apportare eventualmente delle correzioni;
- **codifica**, ovvero di trasformazione dei valori delle variabili in codici numerici o alfanumerici al fine di rendere le variabili più facilmente trattabili in sede di analisi statistica, o di riclassificare le variabili osservate.

## Esempio

Sono stati raccolti i dati relativi alle caratteristiche di 56 ponti:

Le colonne contengono le variabili

Le righe (record)  
rappresentano le  
unità statistiche

IDc	STRADA	TIPO_PONTE	TIPO_MATERIALE_IMPALCATO	OSTACOLO	SCHEMA_STRUTTURALE_IMPALCATO	NUMERO_CAMPATE	LUNGHEZZA_METRI	LARGHEZZA_IMPALCATO	NUMERO_CORSIE
1	A22	Ponte	Calcestruzzo armato	Rete principale	Semplice appoggio	3	45	5	1
2	A22	Ponte	Calcestruzzo armato	Rete principale	Semplice appoggio	3	45	8	2
...	...	...	...	...	...	...	...	...	...
7	A22	Ponte	Acciaio	Rete principale	Arco	2	35	8	2
8	A22	Ponte	Calcestruzzo armato	Multi-ostacolo	Semplice appoggio	5	150	24	4
...	...	...	...	...	...	...	...	...	...
54	SP19	Tombino	Calcestruzzo armato	Fiume	Scatolare	1			2
55	SP14	Tombino	Calcestruzzo armato	Fiume	Scatolare	1	5	8	2
56	SP14	Ponte	Calcestruzzo armato	Fiume	Semplice appoggio	1	30	7	2

Nella base dati non devono esserci righe o colonne completamente vuote. I **dati mancanti** devono essere codificati in modo appropriato (es. cella vuota). La eventuale **codifica** delle variabili deve essere uguale per tutte le unità statistiche. Le **classificazioni** adottate (ovvero l'insieme dei valori assunti dalle variabili) deve essere uguale per tutte le unità statistiche.

ID	STRADA	TIPO_PONTE	TIPO_MATERIALE_IMPALCATO	OSTACOLO	SCHEMA_STRUTTURALE_IMPALCATO	NUMERO_CAMPATE	LUNGHEZZA_METRI	LARGHEZZA_IMPALCATO	NUMERO_CORSIE
1	A22	Ponte	Calcestruzzo armato	Rete principale	Semplice appoggio	3	45	5	1
2	A22	Ponte	Calcestruzzo armato	Ostacolo: Rete principale	Semplice appoggio	3 campate	45 e 50 cm	8	2
...	...	...	...	...	...	...	...	...	...
7	A22	Ponte	Acciaio	Rete principale	Arco	2	35	8	2
8	A22	Ponte	Calcestruzzo armato	Multi-ostacolo	Con semplice appoggio	5 camp.	150 metri	24	4
...	...	...	...	...	...	...	...	...	...
54	SP19	Tombino	Calcestruzzo armato	Ostacolo Fiume	Schema: Scatolare	1	?		2
55	SP14	Tombino	Calcestruzzo armato	Fiume	Scatolare	1	5	8	2
56	SP14	Ponte	Calcestruzzo armato	Fiume	Semplice appoggio	1	30	7	2



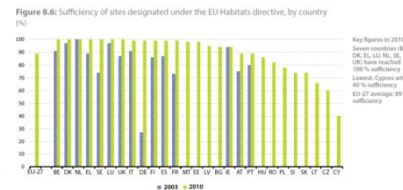
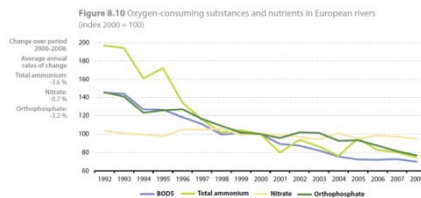
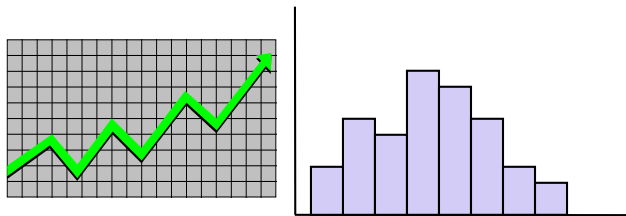
# Le fasi di una indagine statistica

- Analisi statistica e presentazione dei risultati

## Percorso di analisi statistica

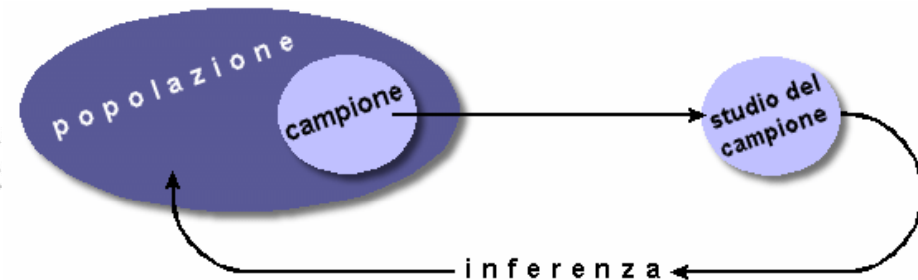
### Statistica descrittiva

Sintesi e descrizione dei dati



### Statistica inferenziale

Trarre conclusioni su intere **popolazioni** a partire da informazioni che si ottengono da **campioni**.



# Le fasi di una indagine statistica

## ▪ Interpretazione e utilizzazione dei risultati della ricerca

I risultati di una indagine statistica consentono di:

- Presentare e descrivere in maniera appropriata dati e informazioni
- Trarre conclusioni su intere popolazioni a partire dalle informazioni che si ottengono da campioni
- Ottenere previsioni affidabili sulle tendenze di fenomeni di interesse
- Migliorare i processi in studio

L'**utilizzo** dei risultati di una ricerca è connessa alle sue modalità di esecuzione e alle regole e ai dettagli definiti durante tutte le fasi dello studio (**modello** e **griglia concettuale**); solo richiamando questi aspetti è possibile dare una **interpretazione** corretta dei risultati.

# Le fasi di una indagine statistica

Per orientarci nell'esecuzione di una indagine statistica, nel sito dell'Istat (www.istat.it) sono consultabili gratuitamente documenti, dispense e altro materiale sotto il menù "Strumenti":

The screenshot shows the Istat website interface. At the top, there is a navigation bar with links for 'Istituto nazionale di statistica', 'Bandi di gara', 'Concorsi', and 'Amministrazione trasparente'. The main header features the Istat logo and a search bar. Below the header, there are tabs for 'Prodotti', 'Strumenti', and 'Informazioni'. The 'Strumenti' tab is active, displaying a list of categories: 'Metodi e software', 'Definizioni e classificazioni', 'Qualità dei dati', 'Cartografia', and 'Dossier'. A large image of various tools (screwdrivers, pliers, etc.) is shown. On the right side, there is a text box stating: 'Su questo sito tutte le informazioni indispensabili per la conoscenza dei metodi e la corretta interpretazione dei risultati delle indagini'. Below this, there are social media icons for Twitter, Facebook, YouTube, and RSS. A sidebar on the right lists the 'Strumenti' categories with sub-items: 'Metodi e software' (Software), 'Definizioni e classificazioni', 'Qualità dei dati' (Linee guida, SIQual, Audit, Riferimenti), 'Cartografia', and 'Dossier'. At the bottom, there are social sharing buttons for Google+, Twitter, Facebook, and LinkedIn.

È possibile recuperare informazioni, software gratuiti, linee guida per le indagini statistiche, ...

Approfondimento

Nel sito dell'Istat (www.istat.it) sotto il menù “Strumenti – Qualità dei dati” inoltre possiamo collegarci al sistema SIQUAL (Sistema Informativo sulla Qualità):

The image shows a screenshot of the Istat website. At the top, there is a navigation bar with the Istat logo and menu items: "Prodotti", "Strumenti", and "Informazioni". A search bar is located to the right of the menu. Below the navigation bar, there is a banner image showing various tools like pens, pencils, and a ruler. To the right of the banner, there is a text box that says "Su questo sito tutte le informazioni".

The main content area is titled "Qualità dei dati" and includes a sub-menu with "Ascolta" and "Strumenti". The "Strumenti" sub-menu is expanded, showing a list of items: "Metodi e Software", "Definizioni e classificazioni", "Qualità dei dati", "Cartografia", and "Dossier". The "Qualità dei dati" item is further expanded to show "Linee guida", "SIQual", "Audit", and "Riferimenti".

On the right side of the page, there is a large graphic for the SIQUAL system. It features the text "SIQual" in a large, stylized font. Below the text, there are four links: "Scelta guidata", "Ricerca multidimensionale", "Elenco completo", and "Documenti". At the bottom of the graphic, there are links for "Webinfo", "Privacy", and "Contatti".

At the bottom of the page, there are three boxes for recent censuses: "15° CENSIMENTO GENERALE DELLA POPOLAZIONE E DELLE ABITAZIONI 2011", "CENSIMENTO DELL'INDUSTRIA E DEI SERVIZI 2011", and "6° Censimento Generale dell'Agricoltura".



Il **Sistema Informativo sulla Qualità** (SIQual) dell'Istat contiene informazioni sulle modalità di esecuzione delle rilevazioni ed elaborazioni condotte dall'Istat su molti fenomeni di interesse in moltissimi settori e sulle attività svolte per garantire la qualità dell'informazione statistica prodotta.

Il sistema è dedicato alla navigazione dei metadati che descrivono le indagini e le sue caratteristiche: contenuto informativo, scomposizione in fasi e operazioni, attività di prevenzione, controllo e valutazione dell'errore.

Nell'area "Documenti" è disponibile la documentazione sia di carattere generale che di interesse specifico per un data indagine.

Se sono disponibili dati on line, attraverso SIQual è possibile consultare tali informazioni.

Tale sistema può essere utile per raccogliere informazioni su un dato argomento di interesse (come **fonte secondaria**), ma può anche essere utilizzato come **guida nel pianificare una data indagine**. Possiamo infatti trovare documentazione su **definizioni** di fenomeni, variabili, **classificazioni**, **questionari**, modalità di rilevazione. Troviamo inoltre le **pubblicazioni** prodotte (annuari, statistiche report, compendi, ...) che possono esserci utili come **guide nei percorsi di analisi statistica e nella scrittura di un report finale** di presentazione dei risultati.

# Il software Minitab per l'analisi statistica dei dati

- Minitab è un pacchetto statistico per applicare un'ampia serie di tecniche statistiche
- Minitab utilizza fogli di lavoro per organizzare i dati (Worksheets)
- Nei fogli di lavoro, Minitab interpreta le colonne come variabili e le righe come unità statistiche.
- Le intersezioni delle colonne e delle righe formano le celle.
- L'insieme di fogli di lavoro, grafici e analisi viene salvato all'interno di un progetto (Project).

## I software statistici sono utilizzati correttamente se:

- Conosciamo come opera il programma
- Conosciamo come organizzare in modo appropriato le informazioni (codifica, presenza di missing, ...)
- Comprendiamo i concetti statistici sottostanti
- Controlliamo la congruità dei risultati evidenziando eventuali errori