

01 – Controllo e Pulizia dei Dati

Giovanni Della Lunga

giovanni.dellalunga@gmail.com

Data Cleaning

- La pulizia dei dati è una delle attività più importanti di Excel. È quella che quotidianamente svolgiamo, talvolta senza nemmeno renderci conto di farla.
- Probabilmente, almeno una volta, vi sarà capitato di ricevere in Excel dei dati esportati da altri software e, nella maggior parte dei casi, questi dati non vengono importati in modo corretto.
- Mi riferisco ai casi in cui i dati contengono spazi extra, celle vuote oppure una formattazione differente (date o numeri interpretati come testo), o ancora casi in cui è necessario manipolare le stringhe di testo.

Data Cleaning

I passaggi di base per la pulizia dei dati sono i seguenti:

- Importare i dati da un'origine dati esterna.
- Creare una copia di backup dei dati originali in una cartella di lavoro separata.
- Verificare che i dati nelle righe e nelle colonne siano in formato tabulare con: dati simili in ogni colonna, tutte le colonne e le righe visibili e nessuna riga vuota all'interno dell'intervallo.
- Eseguire prima le attività che non richiedono la modifica della colonna, ad esempio il controllo ortografico o le operazioni con Trova e sostituisci.

Data Cleaning

Successivamente, eseguire le attività che richiedono la modifica della colonna. I passaggi generali per la modifica di una colonna sono:

- Inserire una nuova colonna (B) accanto alla colonna originale (A) da pulire.
- Aggiungere una formula che trasformerà i dati nella parte superiore della nuova colonna (B).
- Inserire la formula nella nuova colonna (B). In una tabella di Excel una colonna calcolata viene creata automaticamente con i valori inseriti.
- Selezionare la nuova colonna (B), copiarla e incollarla come valori nella nuova colonna (B).
- Rimuovere la colonna originale (A), che converte la nuova colonna da B ad A.

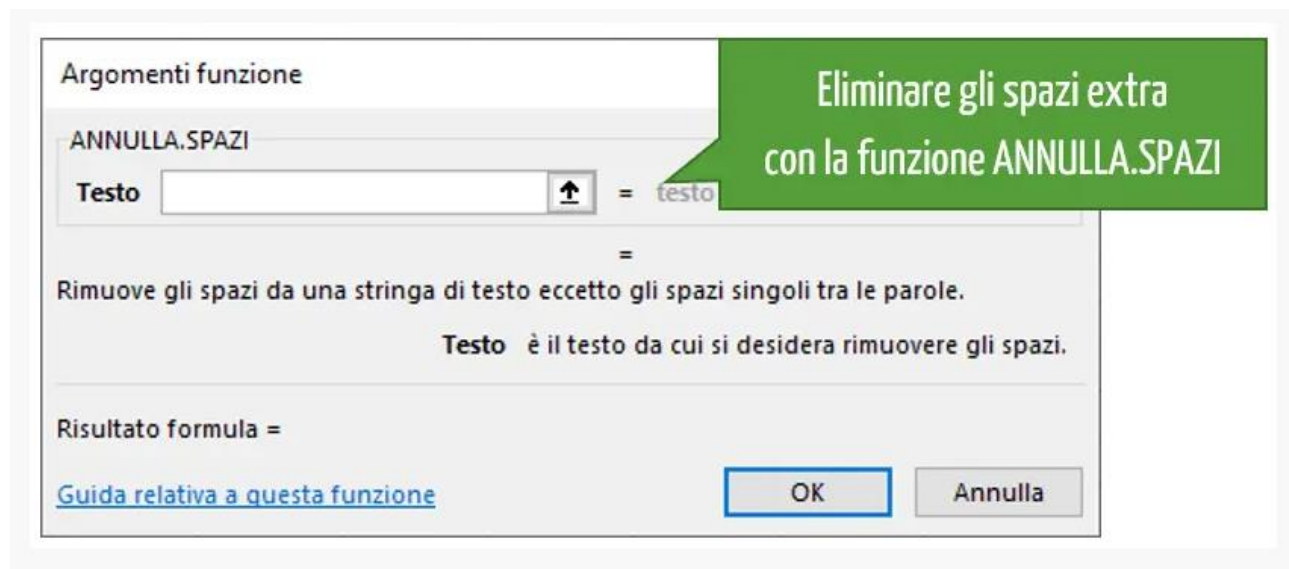
Cancellare la Formattazione

- Cancellare la formattazione delle celle è utile per ripulire i dati e applicare la propria formattazione.
- È possibile eliminare tutta la formattazione attraverso un utile comando.



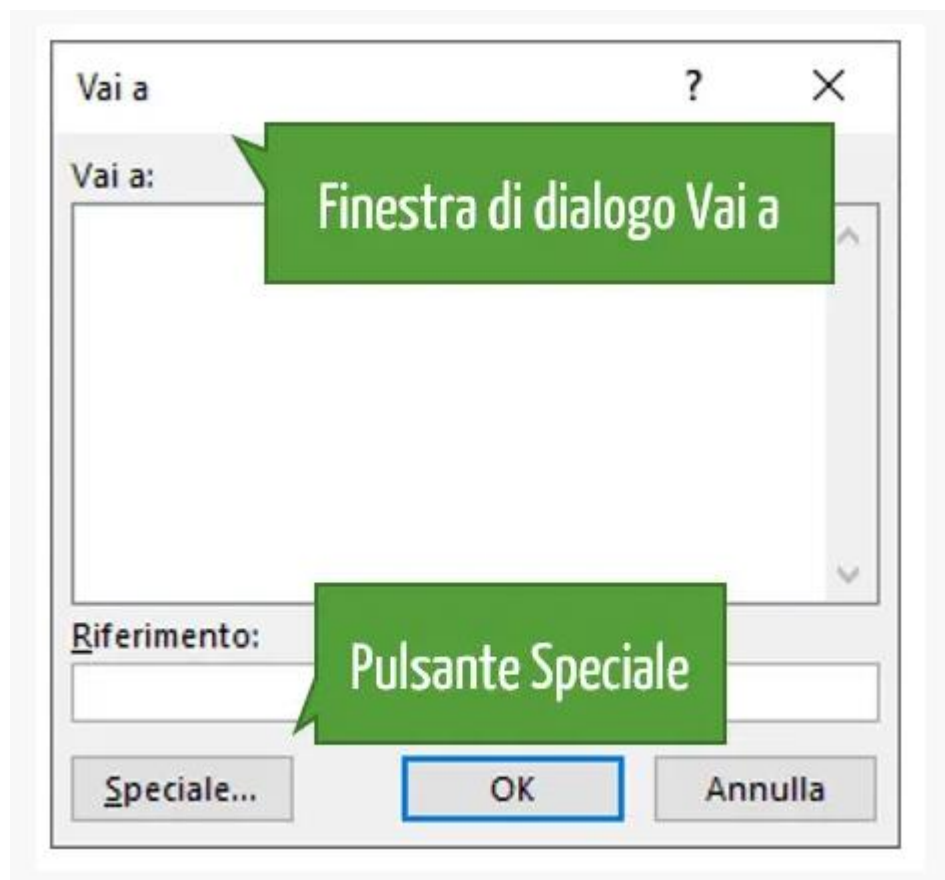
Eliminare Spazi Extra

- Individuare gli spazi extra nelle celle è un'operazione piuttosto difficile. In modo particolare quando si tratta di individuare gli spazi finali.
- Mentre individuare gli spazi extra tra parole o numeri potrebbe risultare più semplice; individuare gli spazi extra finali è spesso, o forse sarebbe meglio dire sempre, un'operazione ardua in quanto non sono visibili.
- Ecco una semplice soluzione per sbarazzarti di questi spazi extra: la funzione ANNULLA.SPAZI.

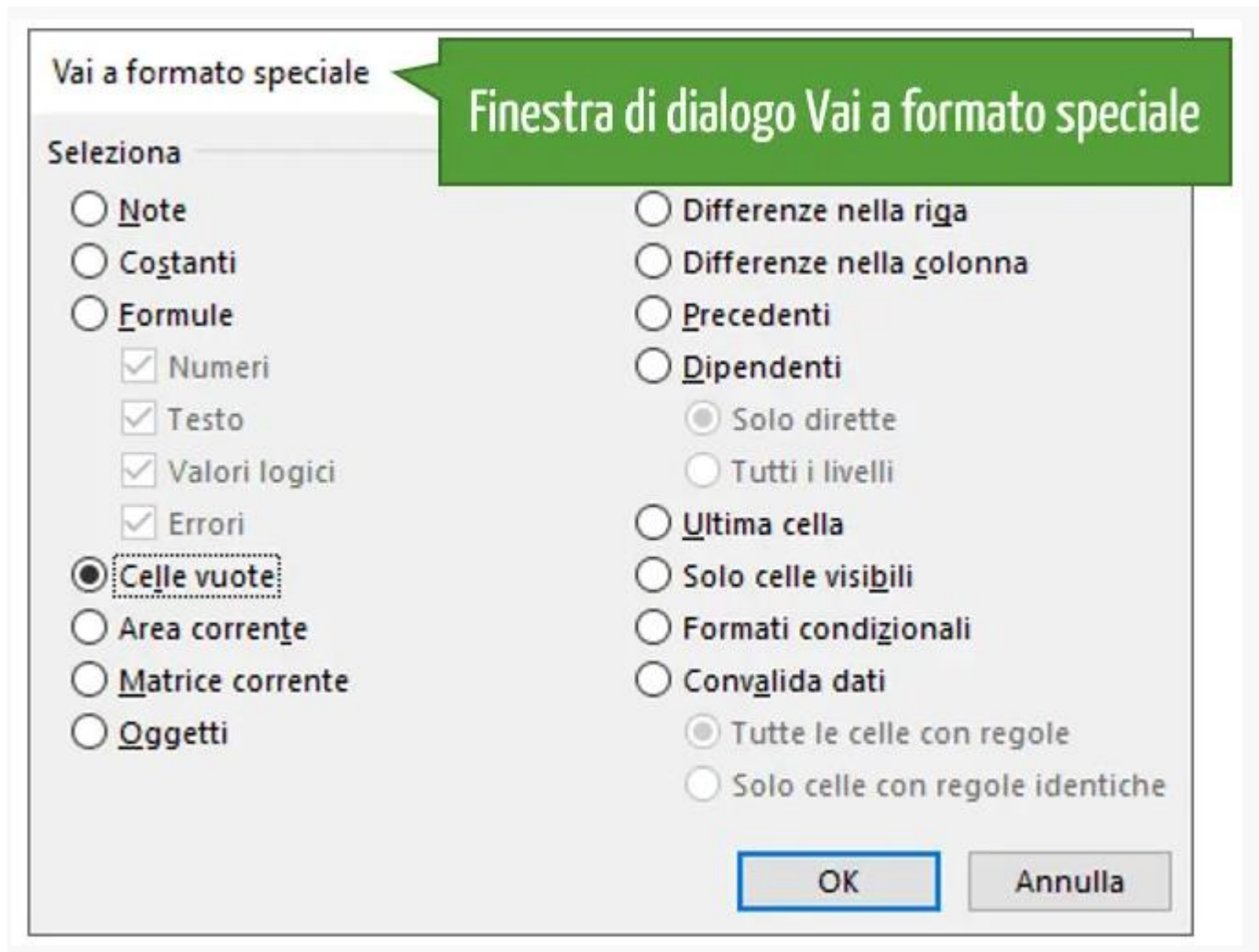


Individua Celle Vuote

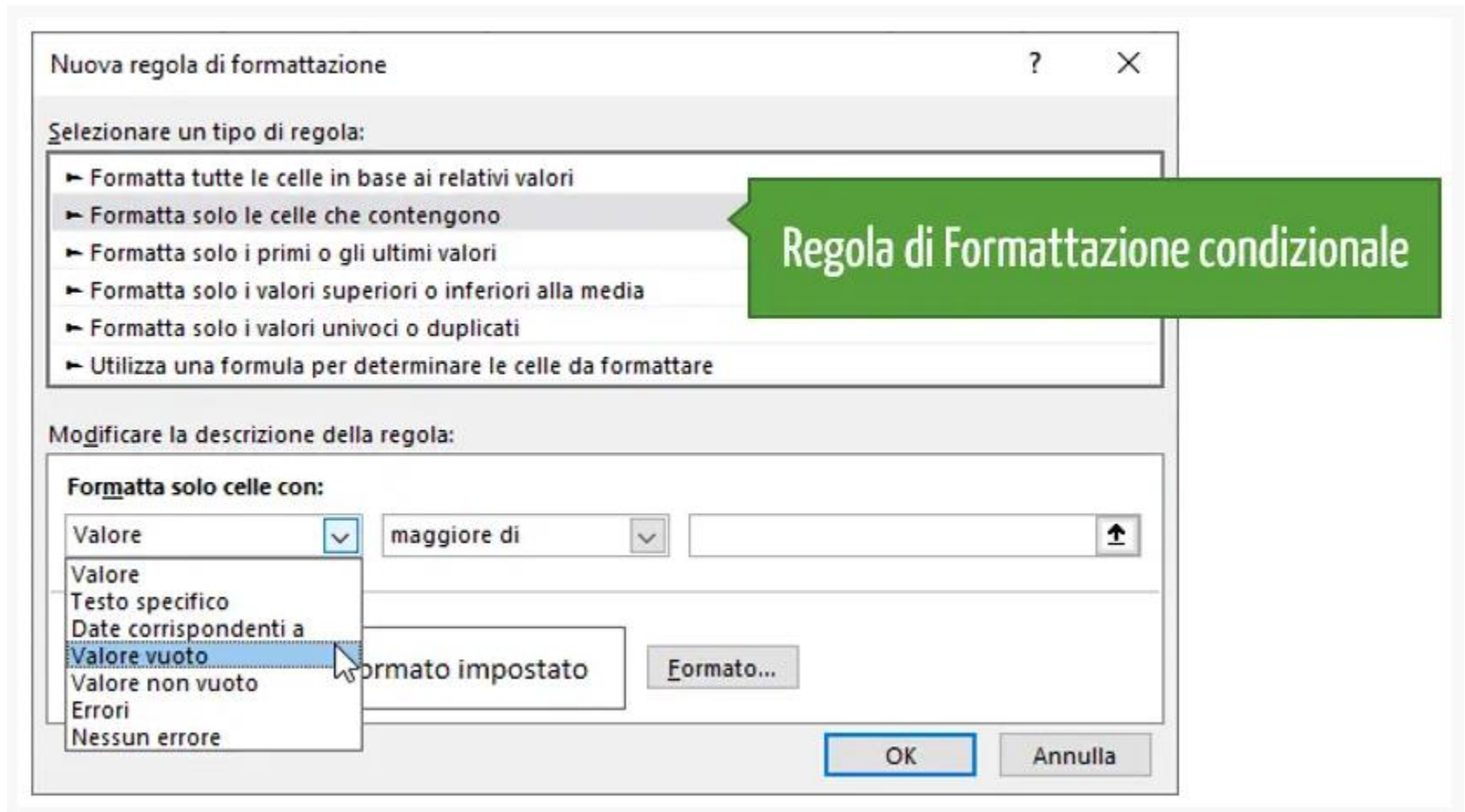
- Selezionare l'intervallo di dati. Premere il tasto funzione F5. A questo punto Excel mostra la finestra di dialogo **Vai a**



Individua Celle Vuote

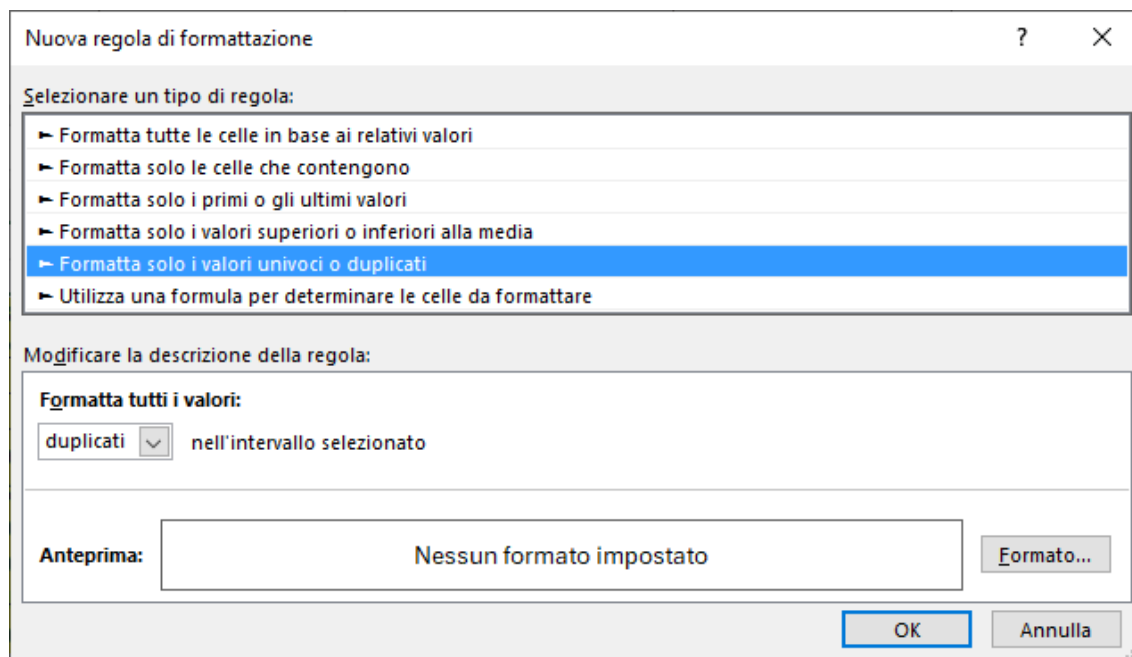


Individua Celle Vuote



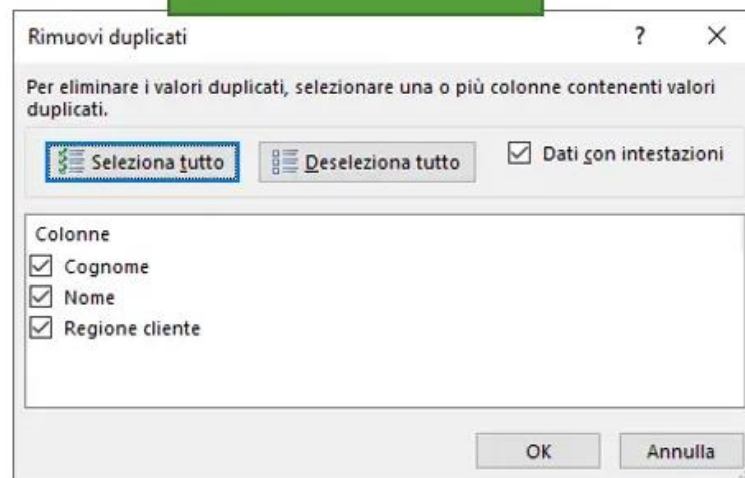
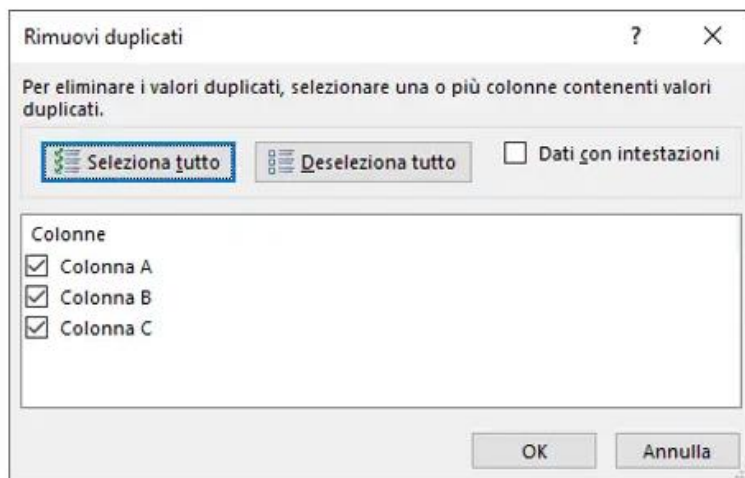
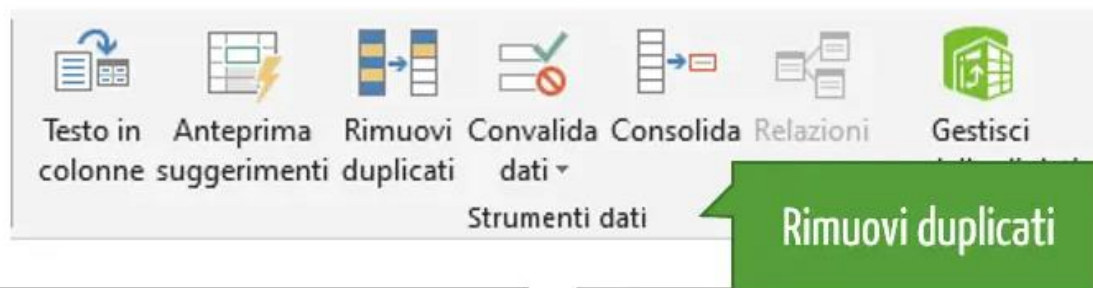
Data Cleaning: Rimuovere Duplicati

- All'interno di un file Excel potrebbero essere presenti dei dati duplicati. Si possono presentare due esigenze:
 - ✓ La prima è evidenziare le celle contenenti dati duplicati.
 - ✓ La seconda è eliminare le celle con i duplicati.
- Il primo caso può essere di nuovo risolto con la formattazione condizionale ...



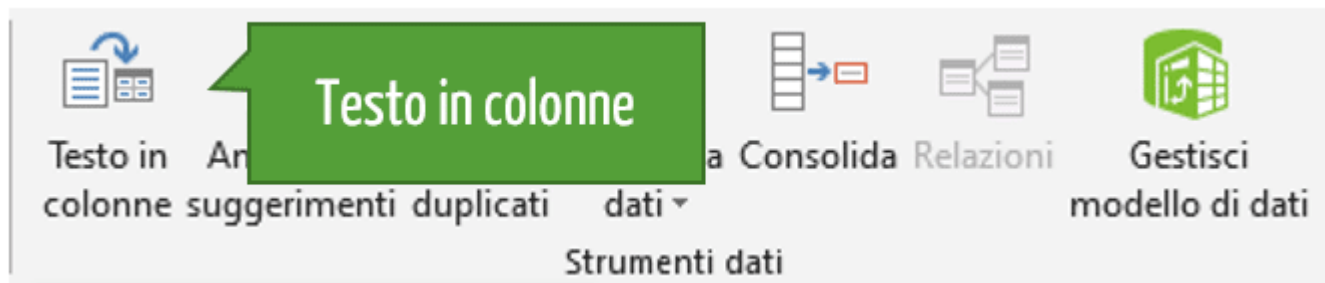
Data Cleaning: Rimuovere Duplicati

- La seconda modalità che puoi utilizzare in presenza di valori duplicati è utilizzare il comando **Rimuovi duplicati**. Questo comando potrai trovarlo all'interno del gruppo **Strumenti dati** nella scheda **Dati**.



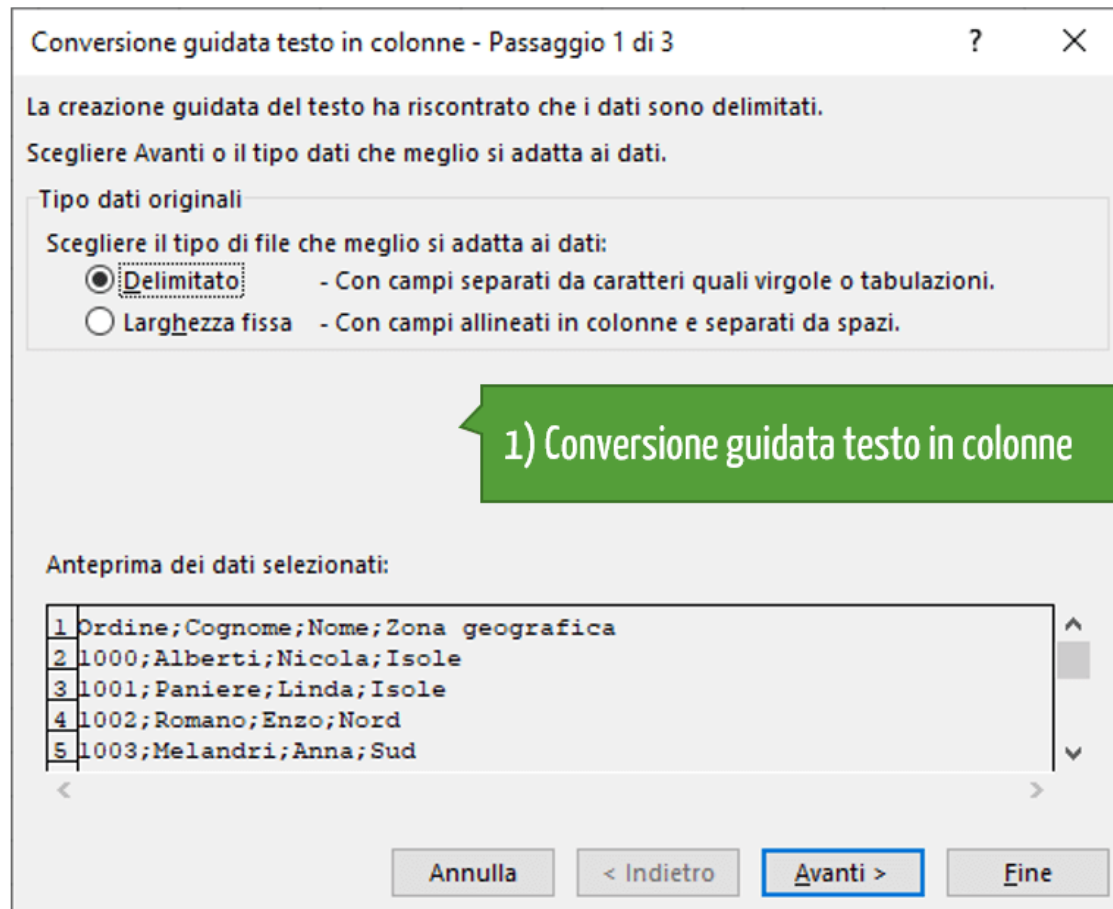
Suddividere il Contenuto di una Cella di Testo

- Importare i dati da un file di testo spesso comporta che i dati vengano inseriti in una cella.
- A questo punto è necessario suddividere il contenuto di una cella di testo in colonne distinte.
- Per compiere questa operazione puoi utilizzare il comando Testo in colonne presente nel gruppo Strumenti dati della scheda Dati.



Suddividere il Contenuto di una Cella di Testo

- Innanzitutto, seleziona le celle contenenti il testo da suddividere.
- Fai un clic sulla scheda Dati e fai un clic sul comando Testo in colonne.
- Excel mostrerà una finestra di dialogo con una procedura guidata.
- Nel Passaggio 1 di 3, seleziona il tipo di dati: Delimitato o Larghezza fissa.



Suddividere il Contenuto di una Cella di Testo

- Nel Passaggio 2 di 3, seleziona il carattere che separa i dati (Delimitatore).
- Qualora il carattere non fosse presente tra i delimitatori elencati, potrai utilizzare l'opzione Altro.

Conversione guidata testo in colonne - Passaggio 1 di 3

La creazione guidata del testo ha riscontrato che i dati sono delimitati.
Scegliere Avanti o il tipo dati che meglio si adatta ai dati.

Tipo dati originali

Scegliere il tipo di file che meglio si adatta ai dati:

☒ Delimitato - Con campi separati da caratteri quali virgole o tabulazioni.

☐ Larghezza fissa - Con campi allineati in colonne e separati da spazi.

1) Conversione guidata testo in colonne

Anteprima dei dati selezionati:

1	Ordine;Cognome;Nome;Zona geografica
2	1000;Alberti;Nicola;Isole
3	1001;Paniere;Linda;Isole
4	1002;Romano;Enzo;Nord
5	1003;Melandri;Anna;Sud

< >

Annulla < Indietro **Avanti >** Fine

Suddividere il Contenuto di una Cella di Testo

- Infine, nel Passaggio 3 di 3, seleziona il formato dei dati e specifica la cella di destinazione.

Conversione guidata testo in colonne - Passaggio 3 di 3

Qui è possibile selezionare ciascuna colonna ed impostare il Formato dati.

Formato dati per colonna

☒ Generale
☐ Testo
☐ Data: GMA
☐ Non importare colonna (salta)

'Generale' converte valori numerici in numeri, valori data in date e i restanti valori in testo.

Avanzate...

Destinazione: =F\$1

Anteprima dati

Standard	Standard	Standard	Standard
Ordine	Cognome	Nome	Zona geografica
1000	Alberti	Nicola	Isole
1001	Paniere	Linda	Isole
1002	Romano	Enzo	Nord
1003	Melandri	Anna	Sud

3) Conversione guidata testo in colonne

Annulla < Indietro Avanti > Fine

Conversione di Testo

- Dopo aver importato i dati da file di testo, capita spesso che alcuni parti del testo non siano corrette ad esempio i nomi ed i cognomi non contengono i caratteri iniziali in maiuscolo. Oppure potresti avere la necessità di scrivere tutto il testo in caratteri maiuscoli. In tutti questi casi puoi utilizzare le seguenti funzioni di testo:
 - ✓ **MAIUSC**
 - ✓ **MINUSC**
 - ✓ **INIZ**
- La funzione MAIUSC di Excel (in inglese nota come funzione UPPER) converte una stringa di testo da minuscolo a maiuscolo.
- La funzione MINUSC di Excel (in inglese nota come funzione LOWER), converte le lettere contenute in una stringa di testo da maiuscole a minuscole.
- La funzione MAIUSC.INIZ di Excel (in inglese nota come funzione PROPER), converte in maiuscolo la prima lettera di ciascuna parola in una stringa di testo e converte le altre in minuscolo.

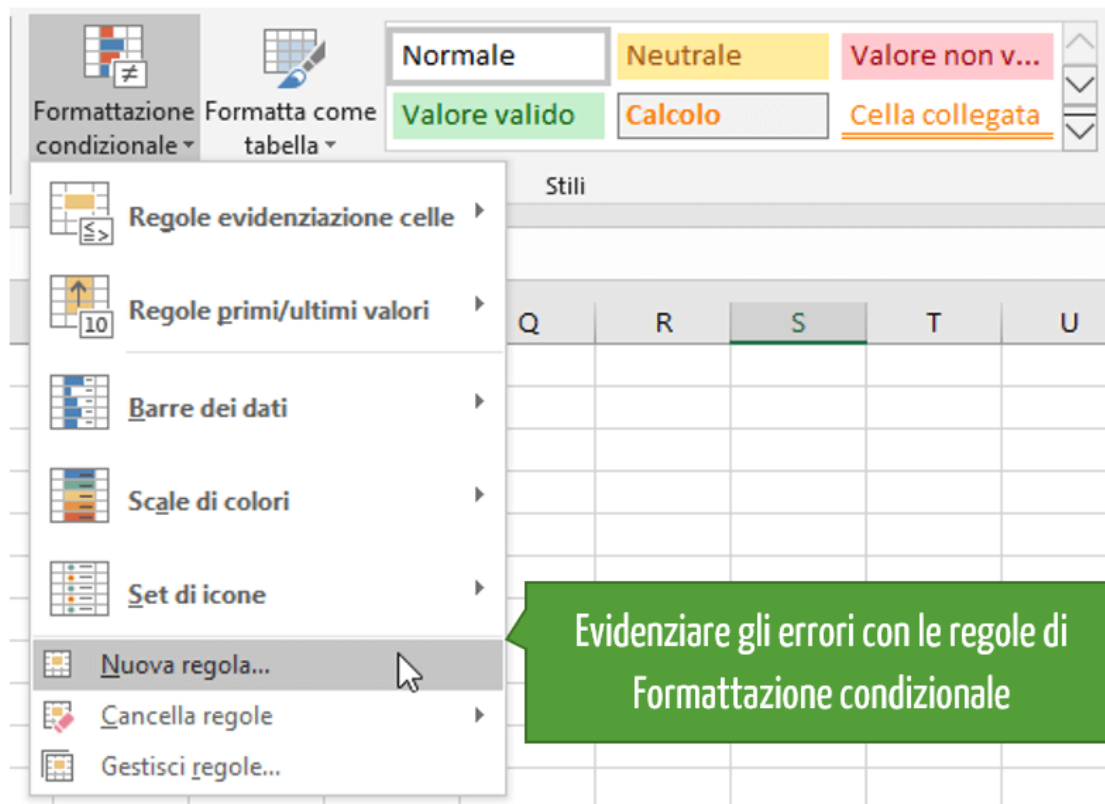
Evidenziare gli Errori

- Il controllo degli errori è **fondamentale**.
- Per individuare gli eventuali errori presenti nel tuo foglio di lavoro puoi utilizzare due modalità.
 - ✓ Le regole di Formattazione condizionale.
 - ✓ Il comando Vai a formato speciale.

Evidenziare gli Errori

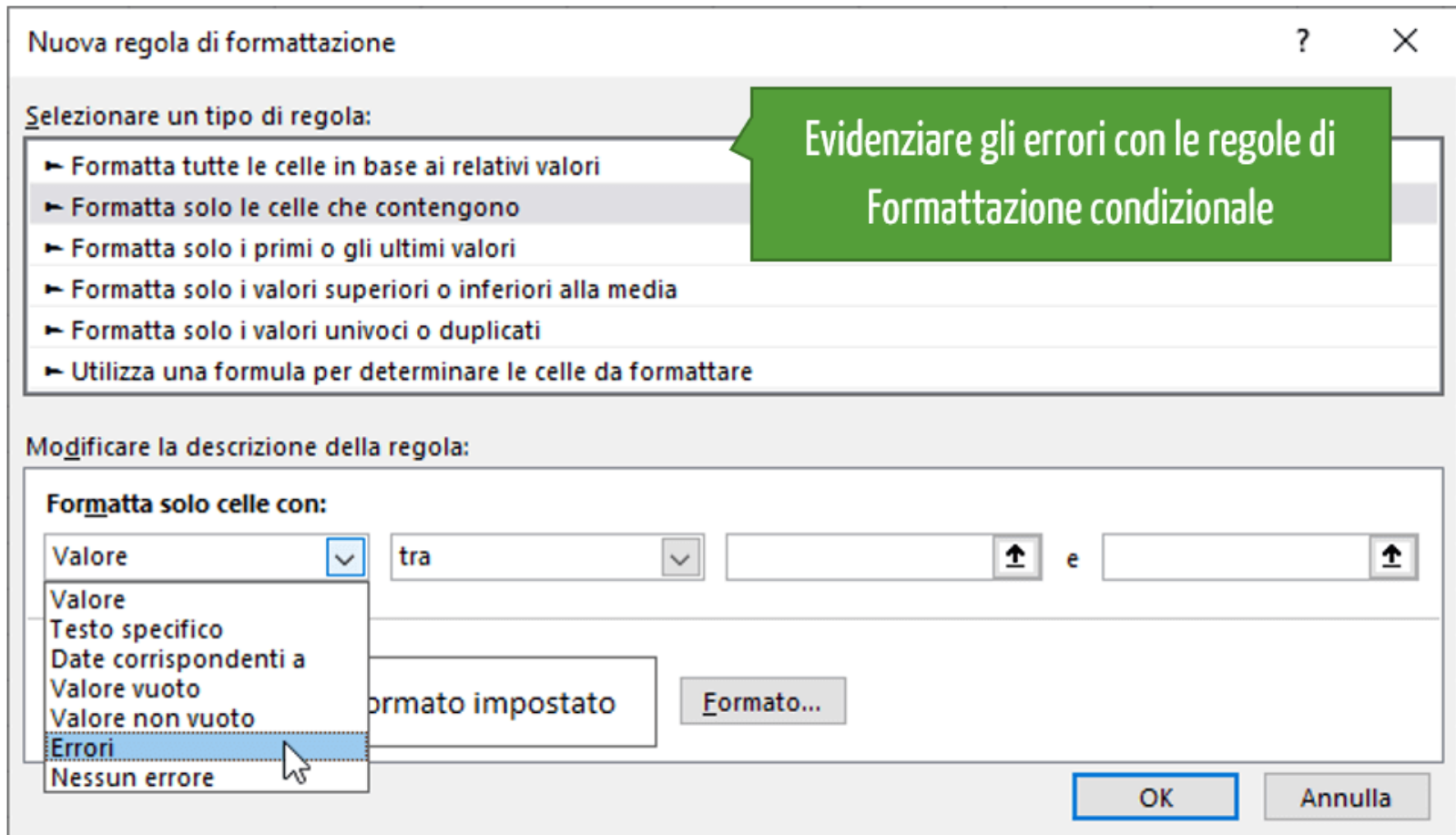
Evidenziare gli errori con le regole di Formattazione condizionale

- Seleziona i dati e vai sulla scheda Home.
- Fai un clic su Formattazione condizionale e seleziona Nuova regola.



Evidenziare gli Errori

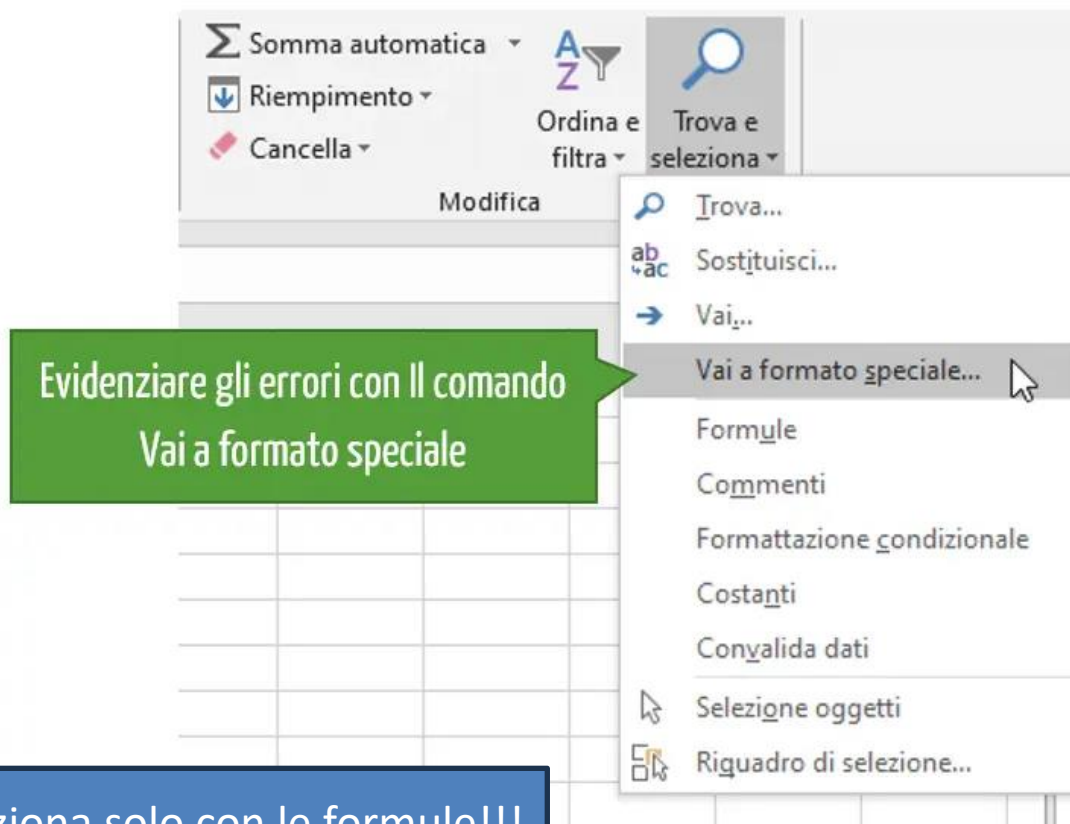
- Nella finestra di dialogo **Nuova regola di formattazione**, seleziona la regola **Formatta solo le celle che contengono**.
- Poi seleziona **Errori** dal menu a discesa.



Evidenziare gli Errori

Evidenziare gli errori con il comando Vai a formato speciale

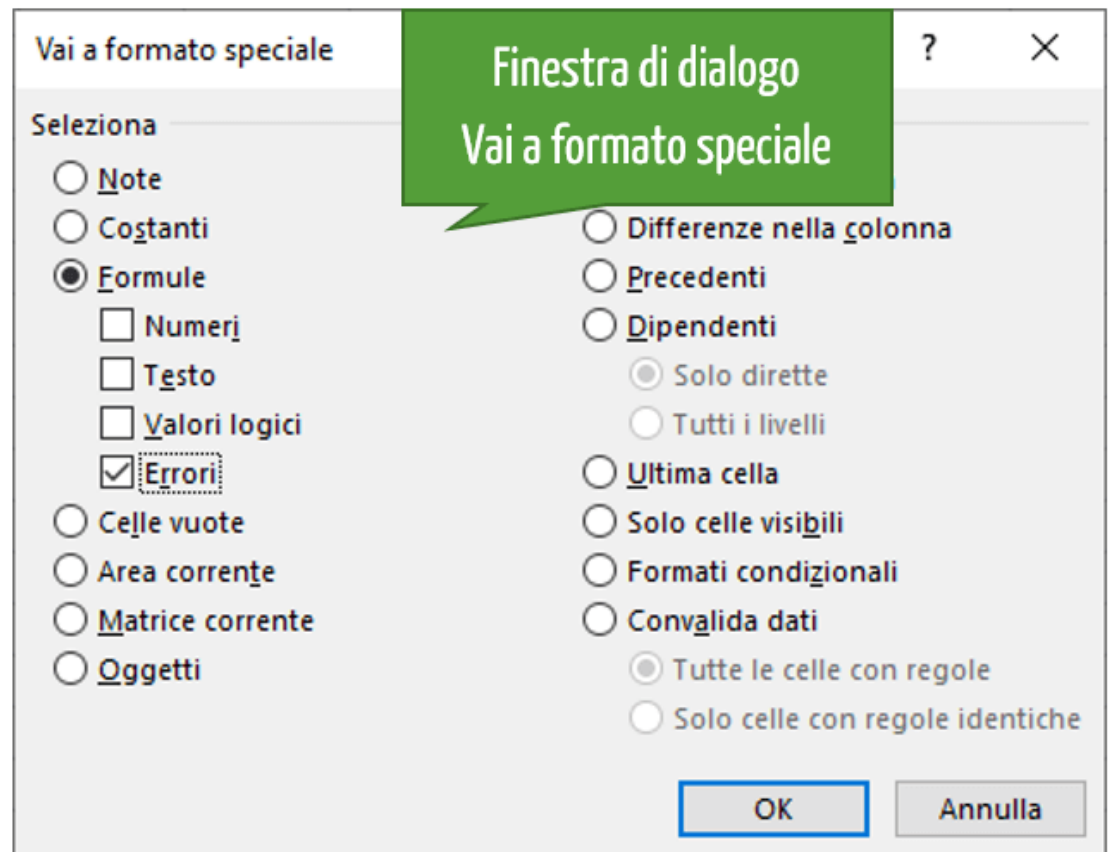
- La seconda modalità per evidenziare gli errori in Excel è utilizzare il comando Vai a formato speciale presente nel gruppo Modifica della scheda Home.



Nota: questo metodo funziona solo con le formule!!!

Evidenziare gli Errori

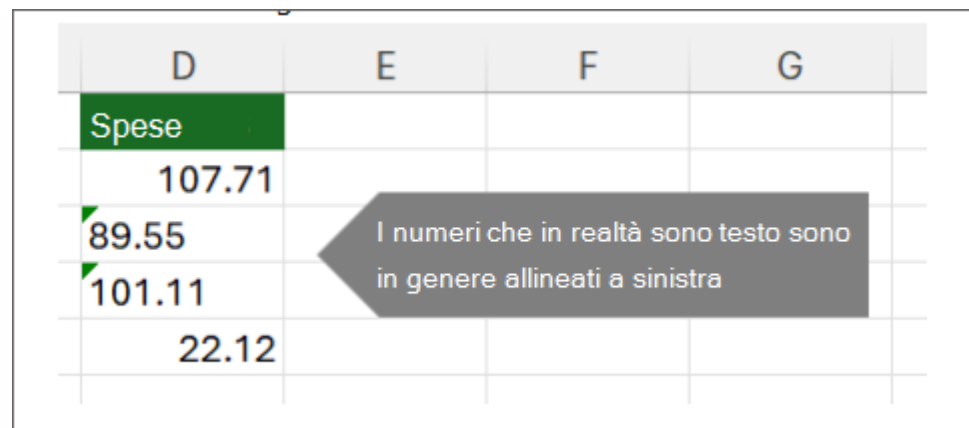
- Dopo aver selezionato i dati fai un clic sul comando Vai a formato speciale. In alternativa, premi il tasto funzione F5.
- In entrambi i casi verrà mostrata la relativa finestra di dialogo.
- Seleziona l'opzione Formule e deselecta tutte le opzioni tranne Errori. Infine, fai un clic su OK.



Correzione di Numeri

Esistono due principali problemi con i numeri che potrebbero richiedere la pulizia dei dati:

- il numero è stato inavvertitamente importato come testo e
 - il segno negativo deve essere modificato nello standard per l'organizzazione.
-
- I numeri archiviati come testo possono causare risultati imprevisti, ad esempio una formula non calcolata visualizzata al posto di un risultato.



The image shows a screenshot of an Excel spreadsheet with columns D, E, F, and G. Column D contains the following values: 'Spese' (highlighted in green), '107.71', '89.55', '101.11', and '22.12'. The numbers '89.55' and '101.11' are preceded by a small green triangle, indicating they are stored as text. A grey callout box with a white arrow points to these two cells, containing the text: 'I numeri che in realtà sono testo sono in genere allineati a sinistra'.

D	E	F	G
Spese			
107.71			
89.55			
101.11			
22.12			

Correzione di Numeri

Altre funzioni utili sono le seguenti ...

VALUTA	Converte un numero in formato testo e applica un simbolo di valuta.
TESTO	Converte un valore di testo in un formato numerico specifico.
FISSO	Arrotonda un numero al numero specificato di decimali, lo formatta in formato decimale usando un punto e le virgole e restituisce il risultato in forma di testo.
VALORE	Converte una stringa di testo rappresentante un numero nel numero corrispondente.

Correzione di Date e Ore

- I formati di data disponibili sono numerosi e possono essere confusi con codici di parti numerati o altre stringhe contenenti barre o trattini, quindi spesso le date e le ore devono essere convertite e riformattate.
- In alcuni casi, le date possono essere formattate e archiviate nelle celle come testo. È ad esempio possibile che sia stata immessa una data in una cella formattata come testo o che i dati siano stati importati o incollati come testo da un'origine dati esterna.
- Le date formattate come testo sono allineate a sinistra in una cella anziché a destra. Quando il controllo degli errori è abilitato, anche le date di testo con anni a due cifre potrebbero essere contrassegnate da un indicatore di errore.

Strumenti per l'Analisi del Testo

- La maggior dei testi sono importati in Excel come parte delle basi dati, mentre i testi vengono inseriti e gestiti da noi prevalentemente durante la creazione o la revisione dei fogli Excel. Raramente c'è la necessità di gestire o manipolare testi durante l'uso ordinario di un foglio di lavoro, se non per aggiunge note nelle pagine di elaborazione, oppure per compilare un report o scrivere la mail con cui condividiamo i risultati dell'elaborazione.
- Quando allora abbiamo bisogno di usare funzioni o altri strumenti per manipolare o elaborare testi?
- Di solito per ragioni precise, come in presenza di **basi dati inaffidabili, o testi "sporchi" da ripulire o da cui estrarre stringhe**, oppure in presenza di errori di formato, o per la necessità di gestire o creare chiavi univoche, e altro ancora.

Strumenti per l'Analisi del Testo

- Quali sono le funzioni Excel di testo?
- Sono una trentina di funzioni che offrono scorciatoie comode ed efficaci per agire sui testi in tutti i modi possibili, dalla rimozione di caratteri specifici (es: spazi o caratteri non stampabili), alla concatenazione di testi diversi, dalla ricerca di testi precisi alla loro estrazione, dalle conversioni al controllo, eccetera.
- Vediamo subito alcuni esempi nel foglio excel relativo a questa sezione...