



Analisi dei dati

Lezione 8

Statistica Descrittiva

Statistica Descrittiva

- La statistica descrittiva mette a disposizione il calcolo di indicatori sintetici che individuano, con un singolo valore, proprietà statistiche di un campione/popolazione rispetto ad una sua variabile/attributo.
- In particolare:
 - indicatori di centralità: media aritmetica, moda, mediana;
 - indicatori di variabilità: varianza, deviazione standard;

Media Aritmetica

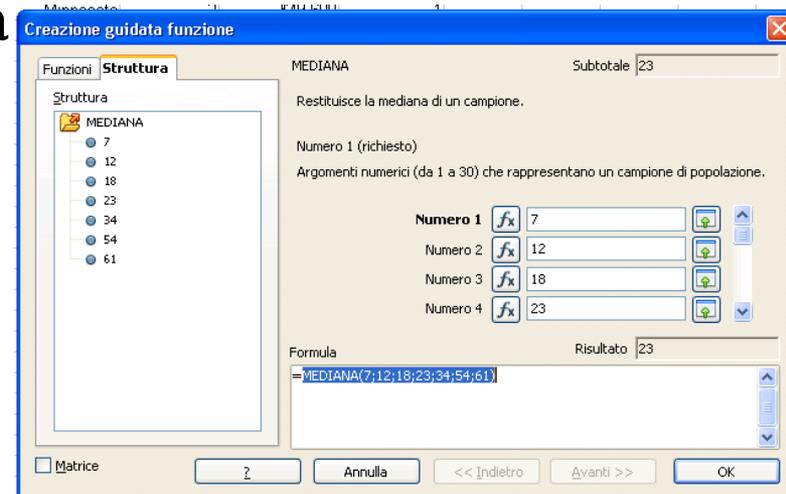
- **MEDIA(numero 1, numero 2, ...)**
 - Numero 1, numero 2, ... sono da 1 a 30 argomenti di cui calcolare la media aritmetica
 - Gli argomenti debbono essere o numeri o nomi, array, riferimenti a campi che contengono numeri.
 - se un argomento di tipo array o riferimento contiene testo, valori logici o celle vuote, questi valori sono ignorati; sono invece conteggiati gli zero presenti.
- Calcoliamo la media della colonna Salary nel file EnvironmentalPolicy.xls



C	D	E	F	G	H	I	J
State	Children	Salary	Opinion				
Minnesota	1	\$65.400	5				
Texas	2	\$62.000	1				
Ohio	0	\$63.200	3				
Florida	2	\$52.000	5				
California	3	\$81.400	1				
New York	3	\$46.300	5		52263.33		
Minnesota	2	\$49.800	1				
New York	1	\$48.100	4				
Virginia	0	\$58.100	2				

Mediana

- Dato un campione di valori *ordinati*
 - $\{X_1, X_1, \dots, X_n\}$
- $MEDIANA = X_{((n+1) \% 2)}$
 - la mediana è quindi il valore centrale del campione.
- In oCalc:
MEDIANA(numero 1, numero 2, ..)
- Robusta: poco influenzata dalla presenza di dati anomali
 - 7 12 18 23 34 54 61
 - Media=29,85
 - Mediana=23



Moda

- la moda è il valore con frequenza più alta nell'insieme delle osservazioni
- In oCalc
 - **MODA(numero 1, numero 2,...)**
- Esempio: calcoliamo la MODA per la colonna “Children” nel file EnvironmentalPolicy.xsl

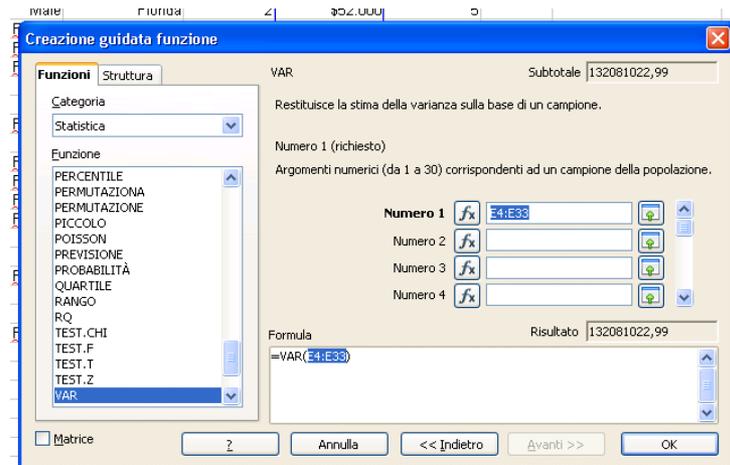
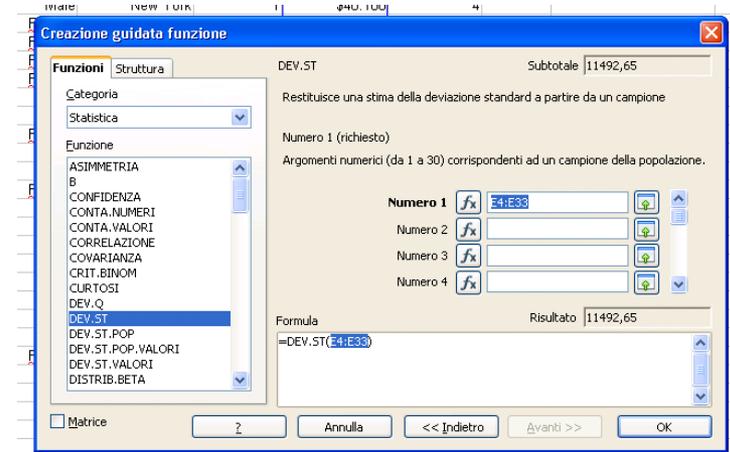
The screenshot shows a spreadsheet with the following data:

State	Children	Salary	Opinion
Minnesota	1	\$65.400	5
Texas	2	\$62.000	1
Ohio	0	\$63.200	3

The 'Creazione guidata funzione' dialog box is open, showing the 'MODA' function. The 'Struttura' tab is selected, and the function is defined as '=MODA(D4:D33)'. The dialog box includes a 'Subtotale' field set to 2, a description of the function, and input fields for 'Numero 1' through 'Numero 4'. The 'Formula' field shows '=MODA(D4:D33)' and the 'Risultato' field shows '2'. The 'OK' button is highlighted.

Varianza e Deviazione Standard

- Deviazione standard = $\text{SQRT}(\text{varianza})$
- In oCalc:
 - $\text{STDEV}(\text{range})$
 - $\text{VAR}(\text{range})$
- Esempio: calcoliamo varianza e deviazione standard per la colonna “Salary” del file EnvironmentalPolicy.xsl



Altre misure descrittive

- **MIN(range)** calcola il valore minimo di un campione memorizzato in range
- **MAX(range)** calcola il valore massimo di un campione
- **RANGE(range)** calcola la differenza tra il valore minimo e massimo, ovvero il campo di variabilità dei valori del campione.