

STATISTICA DESCRITTIVA

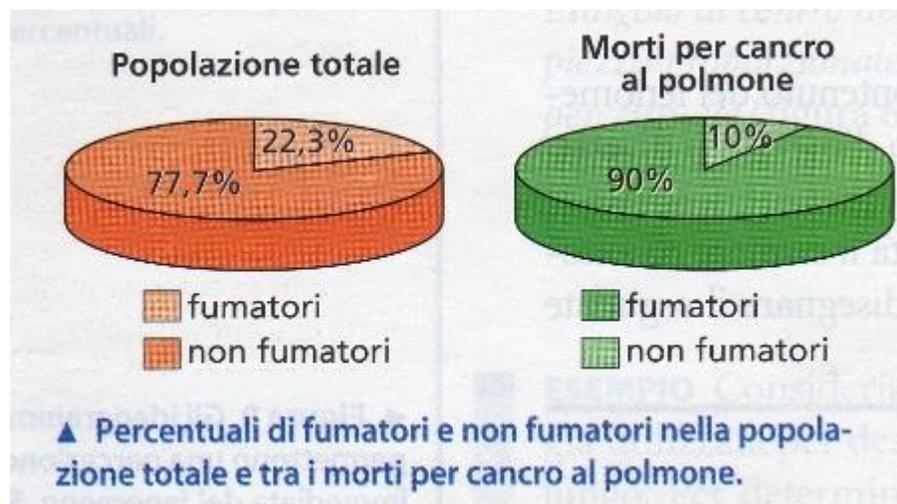
Liceo Scientifico Galileo Ferraris di Torino

Classe 3 E

Aprile 2015

Perché devo studiare statistica?

Ricerca



E' dimostrato che il fumo è un fattore di rischio che favorisce l'insorgenza del cancro al polmone.

Ricerca

Yale-Brown Obsessive Compulsive Scale

- Riferimento a criteri del D.S.M. IV per il disturbo ossessivo compulsivo;
- 10 item: 5 per le ossessioni e 5 per le compulsioni, scala Likert da 0 a 4;
- Si valutano:
 1. Durata e frequenza
 2. Interferenza con la vita sociale e lavorativa
 3. Il disagio soggettivo
 4. Il grado di resistenza e di controllo delle ossessioni e delle compulsioni
- Tempo di somministrazione: 40 minuti circa;
- Monitorizza il decorso e l'efficacia del trattamento.

Analisi di mercato



Descrizione della popolazione



Fare previsioni



Prendere decisioni

		SEPTEMBER 2012				
		Mon 10			Wed 12	
3 participants		9:00 AM	11:00 AM	1:00 PM	9:00 AM	11:00 AM
Volunteer1			✓	✓		✓
Volunteer2		✓		✓	✓	
Volunteer3				✓		
Your name	<input type="checkbox"/>					
		1	1	3	1	1

Statistica descrittiva

L'**obiettivo** è descrivere un gruppo di individui, detto **popolazione**.

- Si sceglie un **campione** della popolazione, in modo che rappresenti l'intero gruppo
- Si effettua l'indagine e si raccolgono i dati
- Si rappresentano graficamente i dati raccolti
- Si elaborano i dati
- Si interpretano i risultati

Statistica induttiva

Dalle interpretazioni dei dati raccolti ed elaborati si vogliono trarre conclusioni sull'intera popolazione.

Si controlla che i risultati trovati possano essere generalizzati all'intera popolazione.

Caratteri

I caratteri si possono classificare in:

- **Qualitativi:** descritti con una parola.
- **Quantitativi:** descritti con un numero. Possono essere a loro volta:
 - **Discreti:** possono assumere solo un numero finito di valori
 - **Continui:** possono assumere infiniti valori in un intervallo reale

Ogni carattere può presentarsi in diverse **modalità**, che descrivono i valori che può assumere.

- Stato civile: carattere qualitativo con due modalità (coniugato/a e libero/a)
- Anno di nascita: carattere quantitativo con più modalità numeriche

Frequenze

Frequenza assoluta: Numero di occorrenze di una data modalità.

Frequenza relativa: $\frac{\textit{frequenza assoluta}}{\textit{numero totale di unità statistiche}}$

Con i caratteri numerici che possono assumere diverse modalità è utile raggruppare le modalità in **classi** e determinare la frequenza di ogni classe.

Frequenza cumulata: somma della frequenza assoluta corrispondente e di tutte le precedenti.

Rappresentazione dei dati

Tabelle

- **Serie statistiche:**

Carattere qualitativo	Frequenze
Modalità 1	Frequenza della modalità 1
Modalità 2	Frequenza della modalità 2

- **Seriazioni statistiche:**

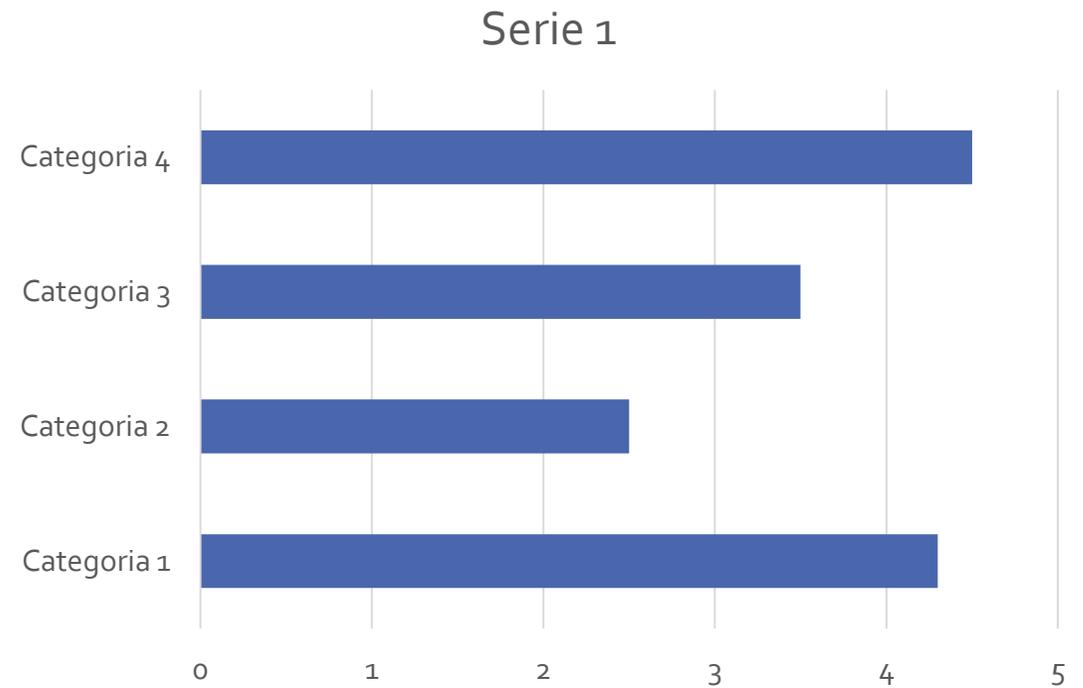
Carattere quantitativo	Frequenze
1	Frequenza della modalità 1
2	Frequenza della modalità 2

- **Tabelle a doppia entrata:** per verificare quando le modalità di due caratteri si presentano contemporaneamente.

Rappresentazione dei dati

Grafici:

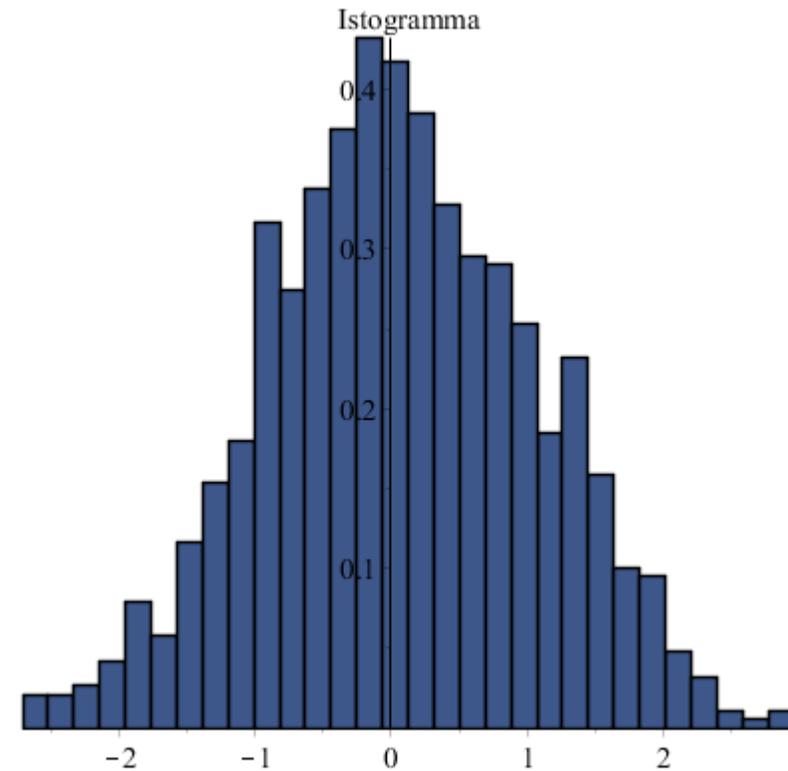
- **Grafico a barre(caratteri qualitativi)**



Rappresentazione dei dati

Grafici:

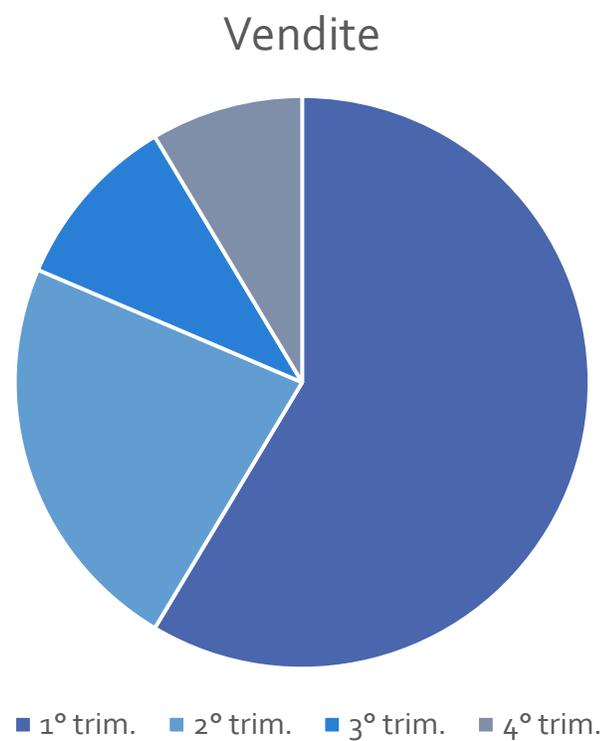
- Istogramma (caratteri quantitativi)



Rappresentazione dei dati

Grafici:

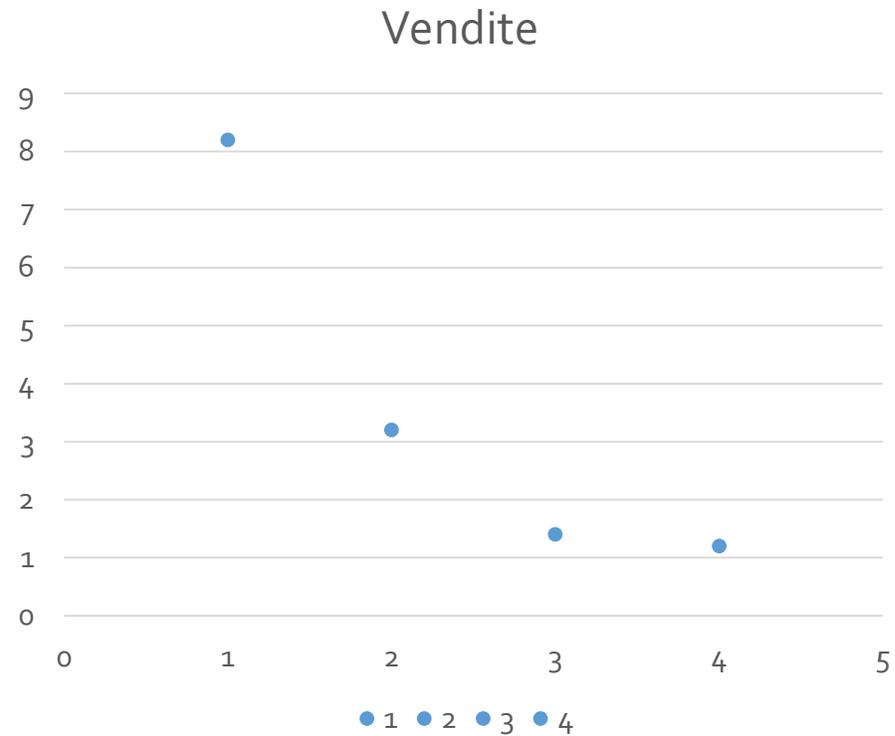
- Areogramma



Rappresentazione dei dati

Grafici:

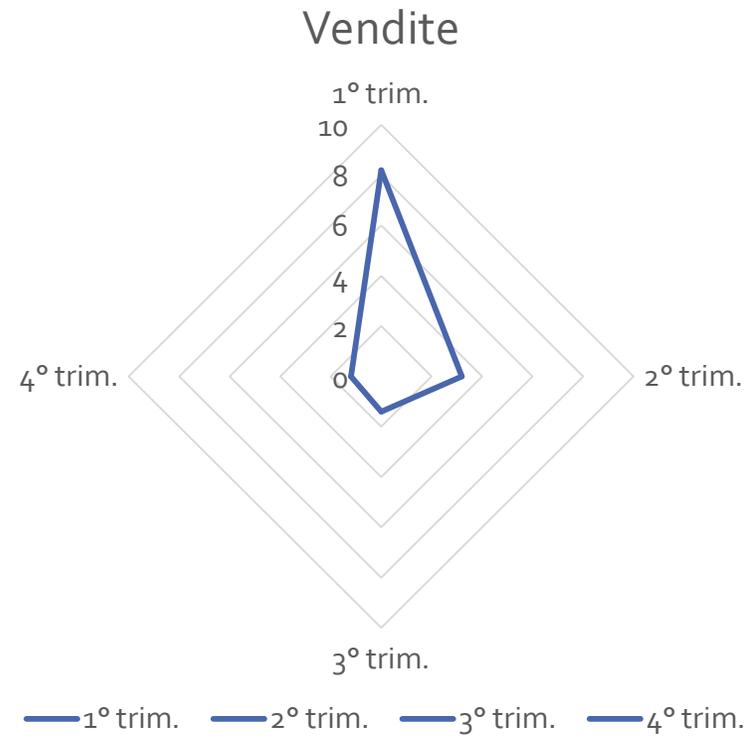
- Diagramma cartesiano



Rappresentazione dei dati

Grafici:

- Radar



Indici di posizione centrale

- **Media aritmetica di n numeri:** rapporto tra somma dei valori e il numero n.

$$M = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

- **Media ponderata di n numeri x_i , ciascuno associato ad un peso p_i :**

$$P = \frac{x_1 p_1 + x_2 p_2 + \cdots + x_n p_n}{p_1 + p_2 + \cdots + p_n}$$

Indici di posizione centrale

- **Media geometrica di n numeri:** radice n-esima del prodotto dei numeri

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

Si applica nei tassi di variazione nel tempo

Anno	Produzione	Variazione (%)
2005	98	
2006	125	27,6
2007	145,5	16,4%
2008	143	-1,7%
2009	165	15,4%

Indici di posizione centrale

- **Mediana di una sequenza ordinata di n numeri:**
 - Il valore centrale, se n è dispari
 - La media dei valori centrali, se n è pari.
- **Moda:** il valore di frequenza massima

Indici di variabilità

- **Campo di variazione:** differenza tra valore massimo e valore minimo di una sequenza di numeri.

Non è molto accurato!

Es: 2, 3, 4, 4, 8, 8, 9, 9, 9, 14

2, 6, 6, 6, 6, 7, 7, 8, 8, 14

- **Scarto semplice medio:** media aritmetica dei valori assoluti degli scarti tra ogni valore e la media

$$S = \frac{|x_1 - M| + |x_2 - M| + \dots + |x_n - M|}{n}$$

Indici di variabilità

- **Varianza:** media degli scarti quadratici dalla media

$$V = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}$$

- **Deviazione standard:** radice quadrata della varianza

$$\sigma = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}}$$

La deviazione standard è l'indice di variabilità più usato.

Proprietà della varianza: la somma degli scarti quadratici rispetto alla media è minima.