

Box plot: una rappresentazione sintetica della distribuzione

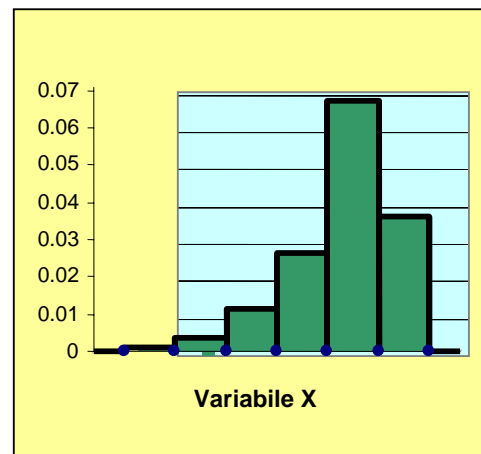
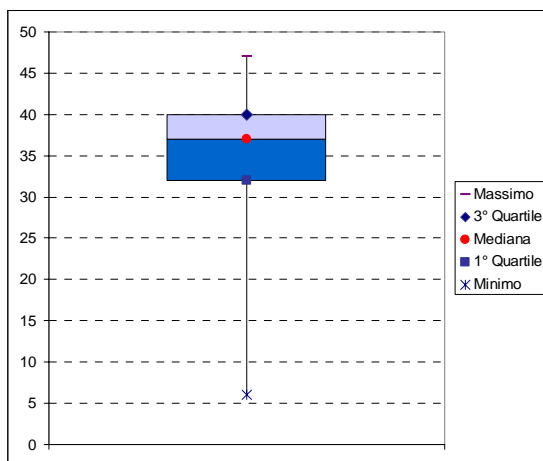
Il box plot o diagramma a scatola e baffi, è un grafico, relativo a caratteri quantitativi - ottenuto a partire dai 5 *numeri di sintesi* [minimo, 1° quartile (Q1), mediana, 3° quartile (Q3), massimo] - che descrive le caratteristiche salienti della distribuzione. Si ottiene riportando su un asse verticale (oppure orizzontale) i 5 numeri di sintesi. La scatola del box plot ha come estremi inferiore e superiore rispettivamente Q1 e Q3. La mediana divide la scatola in due parti. I baffi si ottengono congiungendo Q1 al minimo e Q3 al massimo. In alcuni grafici (ad esempio, quello ottenuto con SPSS) il baffo ha lunghezza pari a 1.5 volte l'altezza della scatola, data dalla distanza tra Q3 e Q1 – detto anche *range interquartile*; ovviamente è inferiore se il massimo valore osservato dista da Q3 meno di 1.5 volte il range interquartile.

Confrontando tra loro le lunghezze dei due baffi (che rappresentano le distanze tra Q1 e il minimo e tra Q3 e il massimo) e le altezze dei due rettangoli che costituiscono la scatola (che rappresentano le distanze tra Q1 e mediana e tra mediana e Q3) si ottengono informazioni sulla simmetria della distribuzione: questa è tanto più simmetrica quanto le lunghezze dei baffi risultano simili tra loro e le altezze dei due rettangoli risultano simili tra loro.

I baffi mettono inoltre in evidenza la presenza di eventuali outliers (osservazioni eccezionali) **[Valori estremi e outliers]**.

Per rappresentare una distribuzione in modo sintetico, il box plot è un'ottima possibilità: con poche informazioni, si riesce a comprendere la sua forma, simmetrica o asimmetrica che sia.

Ad esempio, in questa figura notiamo che il box plot evidenzia efficacemente l'asimmetria della distribuzione del carattere.



Da notare inoltre, che il box plot dà una rappresentazione **univoca** della distribuzione, **a differenza dell'istogramma** che può dare rappresentazioni diverse **a seconda degli estremi delle classi scelte**. Ad esempio, nella pagina di seguito sono riportati 3 istogrammi relativi ad uno stesso carattere, ma ottenuti scegliendo un numero differente di classi di diversa ampiezza.

Il box plot relativo alla distribuzione, però, non varia.

