

Modelli analitici descrittivi

Lezione 2: statistica descrittiva

1 Esercizio 2-1: variabili qualitative ordinali

Una società di marketing ha condotto un'indagine sui consumatori che usano una certa marca di dentifricio. Dopo aver provato un nuovo tipo di dentifricio, alla popolazione dei consumatori indagati è stato chiesto se sarebbero disposti ad acquistare il nuovo prodotto. La risposta è stata valutata sulla seguente scala:

- mai
- poco probabile
- possibile
- probabile
- certo

Il file `Es2-1.txt` riporta le risposte dettagliate. Si illustrino graficamente i dati e si calcolino tutti gli indici di posizione possibili.

Soluzione Si può costruire la tabella delle frequenze assolute con la funzione `CONTA.SE`, e derivarne le frequenze relative e percentuali. Si può quindi disegnare un istogramma e un diagramma a torta della distribuzione di frequenza percentuale. Essendo il carattere qualitativo ordinale, si può definire non solo la moda, ma anche la mediana e gli altri quantili (su cinque valori, i quantili significativi sono molto pochi).

2 Esercizio 2-2: variabili quantitative continue

Il file `Es2-2.txt` riporta i pesi alla nascita di 100 bambini. Si calcolino tutti i possibili indici di posizione e variabilità dei dati. Si raggruppino i dati in una distribuzione di frequenza con un adeguato numero di classi, scegliendo un'ampiezza multipla di 100 per semplicità e si disegnino l'istogramma delle frequenze assolute e il diagramma a linea della frequenza cumulativa assoluta.

Soluzione Qui sono disponibili tutti gli indici, dalla moda alla mediana (con quantili vari) alla media e varianza. È possibile descrivere mediana e quartili interessanti con un diagramma box-plot. Sette classi sono una buona approssimazione (ma anche qualsiasi valore simile).

3 Esercizio 2-3: variabili quantitative

Il file `Es2-3.txt` riporta i risultati di un'indagine sulla distribuzione dei ritardi dei treni a lunga percorrenza nell'arco di due settimane, misurati in minuti. Si calcolino gli indici statistici disponibili, si costruisca una distribuzione di frequenza assoluta,

raggruppando i dati in classi chiuse a destra di uguale ampiezza e se ne disegni il grafico. Si derivino poi le distribuzioni di frequenza relativa e percentuale e si tracci il grafico della distribuzione di frequenza cumulativa assoluta.

Il file `Es2-3b.txt` riporta i risultati di un'indagine analoga condotta l'anno successivo. In base all'analisi e al confronto delle due distribuzioni, discutere se la situazione dei ritardi si possa dire migliorata, peggiorata o invariata.

Soluzione Gli indici statistici possono essere valutati in modo esatto sui singoli dati, oppure in modo approssimato sulle classi. Il primo modo è ovviamente preferibile, ma il secondo potrebbe essere l'unico disponibile se i dati originali non sono di pubblico dominio, e fornisce comunque informazioni approssimate di una certa utilità. Gli indici di variabilità suggeriscono una distribuzione molto ampia dei ritardi, l'indice di asimmetria una pendenza verso destra (ovvia, dato che non ci sono ritardi negativi) e la curtosi un andamento abbastanza "appuntito".

Il file `Es2-3.xls` riporta nel primo foglio di lavoro la soluzione con 8 classi con il corrispondente diagramma box-plot, nel secondo foglio la soluzione per l'indagine condotta l'anno seguente, e nel terzo foglio il loro confronto, da cui emerge un miglioramento della situazione.

4 Esercizio 2-4: variabili quantitative

Il file `Es2-4b.txt` riporta gli stipendi base dei dipendenti di un'azienda, classificati in sei livelli elencati nel file `Es2-4a.txt`. Si calcoli lo stipendio medio. Si costruisca una distribuzione di frequenza assoluta e se ne disegni il grafico. Si derivino poi le distribuzioni di frequenza relativa e percentuale e si tracci il grafico della distribuzione di frequenza cumulativa assoluta.

Soluzione Per costruire la distribuzione di frequenza si può usare la funzione `FREQUENZA`. Dato il basso numero delle classi, indici statistici diversi dalla media o mediana hanno poco significato.