

## Tecniche di analisi dei dati

- Classificazione
  - monodimensionale
  - multidimensionale

14/11/2007

1

---

---

---

---

---

---

---

---

## Tecniche di analisi dei dati

- Calcolo di indici
  - di posizione
    - media (var. cardinali), moda (var. nominali), mediana (var. ordinali)
  - di variabilità
    - varianza, deviazione standard, coefficiente di variazione, differenza interquartile
  - di forma della distribuzione
    - asimmetria  $\Sigma(x^3 f) / \sigma^3$
    - concentraz., appiattimento o Curtosi  $\Sigma(x^4 f) / \sigma^4$

14/11/2007

2

---

---

---

---

---

---

---

---

## Tecniche di analisi dei dati

- Costruzioni di serie
  - trasversali (territoriali, per gruppi sociali, ecc.)
  - storiche (temporali)
- Calcolo di indici
  - di variazione nel tempo
    - assoluta, tendenziale, ecc.
  - di variazione tra i gruppi

14/11/2007

3

---

---

---

---

---

---

---

---

## Tecniche di analisi dei dati

- Tecniche di analisi multivariata
  - analisi di regressione semplice
  - analisi di regressione multipla
  - analisi di regressione logistica e discriminante
  - analisi fattoriale e delle componenti principali

14/11/2007

Fonte: Fabbris L., Statistica multivariata - analisi esplorativa dei dati, McGraw-Hill, 1997

---

---

---

---

---

---

---

---

## Tecniche di analisi dei dati

- Tecniche di analisi multivariata
  - scaling multidimensionale (su almeno due dimensioni, spesso su scale ordinali)
  - analisi di raggruppamento e di segmentazione (cluster analysis, algoritmi CHAID e UNAIDED)
  - analisi congiunta (conjoint analysis)

14/11/2007

Fonte: Fabbris L., Statistica multivariata - analisi esplorativa dei dati, McGraw-Hill, 1997

---

---

---

---

---

---

---

---

## Tecnica di base: la regressione multipla

- Analisi di dipendenza
- Relazioni tra
  - una singola variabile (criterio) dipendente e
  - diverse variabili (predittori) indipendenti
    - Ad ogni predittore viene assegnato un peso, o coefficiente, che spiega il contributo alla previsione del valore assunto dalla variabile dipendente
- Richiede la specificazione di una variata, iò di una combinazione di variabili indipendenti che meglio predice la variabile dipendente

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p \pm \varepsilon$$

14/11/2007

6

---

---

---

---

---

---

---

---

## Base di partenza:

### ■ Previsione senza predittori

- Calcolo dell'indice
  - Media
- Calcolo dell'errore della previsione
  - Sommatoria degli errori al quadrato

14/11/2007

7

---

---

---

---

---

---

---

---

## Passo avanti:

- Previsione con un predittore
  - Tentativo di ridurre l'errore
- Valutazione dell'associazione (correlazione) con la variabile dipendente
  - Due variabili sono correlate se un cambiamento nella prima variabile corrisponde ad un cambiamento nella seconda
  - Valutazione della matrice di correlazione
  - Ruolo dell'intercetta
  - Minimizzazione dei **residui** con il metodo dei minimi quadrati
- Definizione di un intervallo di confidenza della stima
  - Errore standard della stima (o deviazione standard degli scarti) = somma degli scarti quadratici / (osservazioni - 2)
  - Calcolo del numero di ESS da aggiungere e togliere al valore stimato in base al livello di confidenza prescelto

14/11/2007

8

---

---

---

---

---

---

---

---

## Passo avanti:

- Valutazione del modello di regressione semplice e del suo errore nella predizione
  - SST: Somma degli scarti quadratici totali (tra valore vero e media)
    - =
  - SSE: Somma degli scarti quadratici tra stima e valore vero (errori)
    - +
  - SSR: Somma degli scarti quadratici tra stima e valore medio (regressione)
- Calcolo del coefficiente di determinazione ( $R^2$ )
  - $SSR/SST$

14/11/2007

9

---

---

---

---

---

---

---

---

## Modello di regressione multipla:

- Previsione con diversi predittori
- Problema della multicollinearità e della indipendenza tra predittori
- Valutazione del contributo di variabili aggiuntive al miglioramento del potere predittivo del modello
  - Valutazione del cambiamento del  $R^2$
  - Decisione di un criterio di interruzione per l'aggiunta di variabili

14/11/2007

10

---

---

---

---

---

---

---

---

## Esercitazione pratica:

- Individuare un fenomeno descritto con una variabile continua
- Valutare quali elementi possono incidere su quel fenomeno
- Misurare il fenomeno e gli elementi che si ritengono ad esso associati con un certo numero di osservazioni
- Valutare la correttezza delle ipotesi utilizzando un foglio di calcolo

14/11/2007

11

---

---

---

---

---

---

---

---