Sommario

Presentazione dell'edizione italiana

*Di Fabio Corradi**

Statistica per data scientist di Alan Agresti e Maria Kateri è, a mio parere, un testo che mancava nella vasta offerta di libri per l'università. Ciò che lo distingue, come testo per un primo corso di statistica, è l'indice, che comprende argomenti standard ma anche argomenti mai presentati a questo livello, quali un bellissimo capitolo sui modelli lineari generalizzati e un'introduzione alla classificazione seguendo sia l'approccio modellistico sia quello algoritmico. Di più: quasi tutti i temi vengono trattati sia con l'approccio frequentista sia con quello Bayesiano. L'approccio frequentista è illustrato in modo onesto e approfondito. L'approccio Bayesiano è ben illustrato ma forse poco valorizzato nei commenti dei risultati. Il tutto lascia al docente l'opportunità di operare i dovuti confronti fra le due scuole.

L'approccio didattico è quello di un testo facile con molti esercizi, alcuni fra i quali, il docente li riconoscerà, necessitano di un po' più di teoria di quella illustrata nel testo. Insomma, con un'opportuna scelta nelle assegnazioni si può consolidare la materia esposta a lezione e lasciare gli studenti cimentarsi con qualcosa che, in fase di revisione, farà emergere altri preziosi frammenti di teoria.

Altre piccole perle: diversi accenni all'inferenza causale, e alla selezione di variabili fatta in modo «moderno» ovvero attraverso penalizzazioni. Peccato che in questo caso ci sia solo la versione frequentista e manchi quella Bayesiana delle a priori «spike and slab» che sarebbero ben servite per un ulteriore confronto che potrà sempre essere fatto dal docente integrando un po' il testo. Infine ci sono due Appendici che illustrano l'uso di `R` e `Python` per la statistica. Le due appendici possono essere utilizzate in alternativa, ovvero scegliendo quella più adatta al background degli studenti e ai corsi a cui sono già stati esposti. Nel caso di un/una aspirante data scientist che avesse poca dimestichezza con la programmazione, questo è un viatico ideale perché si procede per esempi articolati sui capitoli del libro, seguendo l'approccio della mano guidata per scrivere le lettere dell'alfabeto.

La platea di studenti a cui il testo è rivolto è quella di corsi di laurea che abbiano avuto possibilmente un primo contatto con il calcolo delle probabilità. L'intero materiale presentato è idoneo per un corso di 9 crediti, ovvero una settantina di ore, essendo sempre possibili semplificazioni che lo riducano a un corso di circa cinquanta ore.

Fabio Corradi

*Fabio Corradi è professore ordinario di Statistica all'università di Firenze.