

# Statistica (c.p.)

## Esercitazioni pratiche

### 1. Previsione del ricorso al legale

Per 1340 sinistri automobilistici con danni a persone si sono rilevate diverse variabili relative all'assicurato o al sinistro stesso compreso l'ammontare del danno.

In alcuni casi l'assicurato ricorre all'assistenza di un avvocato. In questi casi è prevedibile che il contenzioso duri di più, abbia costi legali per l'azienda e che la somma liquidata possa infine rivelarsi più elevata. E' pertanto rilevante capire quali caratteristiche dell'assicurato o del sinistro impattino sulla probabilità di rivolgersi all'avvocato.

- ATTORNEY : se il danneggiato è assistito da un avvocato (yes/no)
- CLMSEX : maschio/femmina
- MARITAL : sposato (M)/libero (S)/vedovo (W)/divorziato (D)
- CLMINSUR : se il guidatore del veicolo danneggiato è assicurato (yes/no)
- SEATBELT : se il danneggiato indossava la cintura di sicurezza/seggolino per bambini (yes/no)
- CLMAGE : età del danneggiato
- AGECLASS : età del danneggiato in classi: (-18] / (18,26] / (26,36] / (36,47] / (47+)
- LOSS : danno subito ( $\times 1000$ )

*Dati: Automobile Bodily Injury Claims (autoBI), Fonte: Frees (2009).*

*<http://instruction.bus.wisc.edu/jffrees/jffreesbooks/Regression%20Modeling/BookWebDec2010/data.html>*

- (a) Determinare un modello che individui quali caratteristiche dell'assicurato o del sinistro favoriscano (o sfavoriscano) il ricorso al legale da parte dell'assicurato.
- (b) Si vuole usare il modello individuato sopra per prevedere il ricorso ad un avvocato per assicurati per i quali si dispone dei valori delle altre variabili. A tal fine si decide di usare il modello trovato sopra adottando il criterio per cui se la probabilità stimata che un assicurato ricorra al legale è maggiore di 0.7 si decide che ricorrerà all'avvocato mentre se è invece inferiore a 0.7 si decide che ne farà a meno. Valutare la capacità del modello ai fini previsivi
  - A tal fine si suggerisce di suddividere casualmente il dataset in due sottoinsiemi, di ottenere sul primo (più numeroso) il modello stimato e poi di verificare sul secondo le capacità previsive.
  - Calcolare un indice della bontà previsiva del modello basato sulla proporzione di casi per i quali la previsione risulta corretta confrontando il valore della previsione con il valore vero del test. Cosa accade se la soglia di 0.7 viene variata?

2. All'indirizzo <http://www.stat.columbia.edu/gelman/arm/examples/risky.behavior/> si trovano i dati relativi all'esercizio che è descritto nel volume di Gelman e Hill (2007) e riportato di seguito

1. Poisson regression: the folder `risky.behavior` contains data from a randomized trial targeting couples at high risk of HIV infection. The intervention provided counseling sessions regarding practices that could reduce their likelihood of contracting HIV. Couples were randomized either to a control group, a group in which just the woman participated, or a group in which both members of the couple participated. One of the outcomes examined after three months was “number of unprotected sex acts.”
  - (a) Model this outcome as a function of treatment assignment using a Poisson regression. Does the model fit well? Is there evidence of overdispersion?
  - (b) Next extend the model to include pre-treatment measures of the outcome and the additional pre-treatment variables included in the dataset. Does the model fit well? Is there evidence of overdispersion?
  - (c) Fit an overdispersed Poisson model. What do you conclude regarding effectiveness of the intervention?
  - (d) These data include responses from both men and women from the participating couples. Does this give you any concern with regard to our modeling assumptions?