Valutazione dei ritardi, correlazione statistica con la puntualità ed attribuzione e validazione dei ritardi in RFI

di Salvatore Carboni

Premessa

Nella valutazione a posteriori dell'andamento della circolazione ferroviaria e dei diversi fattori che possono intervenire a perturbarla, gli analisti hanno tradizionalmente seguito un approccio quanti/qualitativo che da qualche decennio ha potuto avvalersi di un volume di dati sempre più ricco sopratutto per merito della progressiva introduzione di strumenti informatici.
Rimandando alla chiusa di questo articolo una sommaria ricostruzione del quadro normativo e dei collegati strumenti e software dedicati oggi a disposizione, vogliamo portare un semplice esempio di valutazione statistica che, insieme ad altre, possa aiutare ad illuminare le "patologie" della circolazione in modo sempre più approfondito e con un arricchimento degli angoli visuali.

L'indice di correlazione statistica

Facendo appello alle, per molti di noi, lontane fatiche scolastiche sulla statistica, possiamo rammentare che disponendo di una o più serie di dati numerici, è possibile calcolare alcune semplici grandezze che sono diretta conseguenza delle serie stesse.
Queste grandezze sono il risultato di semplici funzioni statistiche che nell'ordine che ci interessa sono:

- la media di una serie di valori che può scriversi $\bar{x} = \frac{1}{n}\sum_{i}^{n} x_i$
- la varianza di una serie di valori, che, aldilà della formula matematica, ci indica quanto i valori stessi sono dispersi, o viceversa addensati, intorno al valore medio
- la deviazione standard, che corrisponde alla radice quadrata della varianza
- la covarianza, che, come è facile intuire, indica se due serie di dati tendano più o meno a variare insieme facendo ipotizzare una correlazione
- infine la formula di correlazione vera e propria che può scriversi

$$\text{Corr}(X,Y) = \frac{COV(X,Y)}{\sigma_x \cdot \sigma_y}$$

e che, come si vede, mette in relazione frazionaria la covarianza tra due serie di dati