

## INDAGINE SUI SERVIZI BIBLIOTECARI PER BAMBINI E RAGAZZI IN TOSCANA

### Appendice – Ricalcolo dei dati

Per ricavare i dati mancanti di due caratteri sono stati calcolati innanzitutto i coefficienti di correlazione lineare fra le variabili in esame.

Tale coefficiente è un importante indice che quantifica il tipo, la forza e la direzione di una relazione tra due variabili quantitative: il tipo di relazione cercata è appunto lineare, la forza (ossia il valore assoluto di tale coefficiente) è compresa tra 0 e 1 e infine la direzione (segno) indica se la  $Y$  (variabile dipendente) varia concordemente di segno con la variazione della  $X$  (variabile indipendente). Di seguito sono mostrate le correlazioni lineari che sussistono fra tutte le possibili coppie di variabili (calcolate sulla base dei soli record che abbiano entrambi i valori delle variabili per cui è valutato tale).

Correlazioni	Pop b/r	Utenti tot	Utenti b/r	Dotazione b/r	Acquisti b/r	Prestiti tot	Prestiti b/r	Ore apertura	Giorni apertura
Pop tot	<b>0,98</b>	<b>0,60</b>	<b>0,62</b>	0,29	0,30	0,53	0,38	0,46	0,30
Pop b/r		0,54	0,56	0,32	0,33	0,55	0,40	0,43	0,31
Utenti tot			<b>0,97</b>	0,13	0,04	0,24	0,16	0,15	0,16
Utenti b/r				0,28	0,14	0,27	0,24	0,14	0,19
Dotazione b/r					<b>0,61</b>	<b>0,53</b>	<b>0,67</b>	0,28	0,21
Acquisti b/r						<b>0,60</b>	<b>0,80</b>	0,36	0,20
Prestiti tot							<b>0,81</b>	<b>0,63</b>	0,35
Prestiti b/r								0,42	0,26
Ore apertura									<b>0,64</b>
Giorni apertura									

Si possono ricavare i dati mancanti di una variabile  $Y$  a partire da un'altra variabile  $X$  se fra di esse sussiste una relazione lineare del tipo:

$$Y = \alpha + \beta \cdot X$$

dove  $\alpha$  e  $\beta$  sono rispettivamente l'intercetta e il coefficiente di regressione.

Si ipotizza dunque una relazione lineare fra i due caratteri maggiormente correlati linearmente (valori in rosso) e conseguentemente si stimano intercetta e coefficiente di regressione (metodo dei minimi quadrati) attraverso i soli record con la coppia di dati disponibile. Mentre per quelli con un solo dato mancante, si ricava quest'ultimo a partire da quello disponibile sostituendone il valore nella retta di regressione stimata. Per i record in cui mancano entrambi i dati oppure per previsione non accettabili si ricorre a successive

relazione lineari sempre tenendo in considerazione la matrice di correlazione (valori in blu e comunque maggiori del 50%).

Le relazioni lineari ipotizzate sono le seguenti:

- 1) Pop b/r in relazione lineare con Popolazione tot (e viceversa) e se mancano entrambi i dati si calcolano sia i valori Pop tot che quelli Pop b/r dalla relazione lineare con Utenti tot; non si possono ricavare entrambe le stime per i record 53 e 129;
- 2) Utente b/r in relazione lineare con Utente tot (e viceversa) e se mancano entrambi i dati si calcolano sia i valori Utente tot che quelli Utenti b/r dalla relazione lineare con Pop tot; non si possono ricavare entrambe le stime per i record 53 e 129;
- 3) Dotazione b/r in relazione lineare con Prestiti b/r e se mancano entrambi i dati si calcolano i valori Dotazione b/r dalla relazione lineare con Acquisti b/r (e ancora da quella con Prestiti tot); non si possono ricavare le stime per i record 53 e 70;
- 4) Acquisti b/r in relazione lineare con Prestiti b/r e se mancano entrambi i dati si calcolano i valori Acquisti b/r dalla relazione lineare con Dotazione b/r (e ancora da quella con Prestiti tot); non si possono ricavare le stime per i record 53 e 70;
- 5) Prestiti b/r in relazione lineare con Prestiti tot (e viceversa) e se mancano entrambi i dati si calcolano i valori Prestiti tot dalla relazione lineare con Ore apertura e i valori Prestiti b/r dalla relazione lineare con Acquisti b/r (e ancora quella con Dotazione b/r); non si possono ricavare la stima di Prestiti tot per il record 53 e le stime di Prestiti b/r per i record 53 e 70 mentre risultano non accettabili e quindi mancanti le stime di Prestiti b/r per i record 62, 74 e 83;
- 6) Ore apertura in relazione lineare con Giorni apertura (e viceversa) e se mancano entrambi i dati si calcolano i valori Ore apertura dalla relazione lineare con Prestiti tot; non si possono ricavare la stima di Ore apertura per il record 53 e le stime di Giorni apertura per i record 4, 5 e 53.

In funzione di tale analisi il campione iniziale di 130 unità è stato ridotto a 127 a causa dell'elevato numero di dati mancanti (17) e non stimabili per le seguenti unità:

- Biblioteca degli Intronati, Siena
- Biblioteca comunale "A. Salucci", Aulla
- Biblioteca Comunale, Monteroni d'Arbia

In conclusione i dati mancanti sono stati ridotti a 5 (3 per Prestiti b/r e 2 per Giorni apertura) costituendo appena lo 0,4% dei dati complessivi.