10. Test di Wilcoxon per campioni indipendenti

Sebbene spesso chiamato test di Mann-Whitney, l'equivalente non parametrico del test t di Student per dati appaiati è dovuto anch'esso a Wilcoxon. Rappresenta l'equivalente non parametrico del test t di Student per campioni indipendenti, e va utilizzato in luogo di questo quando i dati non siano distribuiti in modo gaussiano.

Per i calcoli si procede in questo modo:
→ mettere i dati dei due campioni in una singola lista, badando ad etichettarli in modo che poi possano essere successivamente di nuovo distinti;
→ stabilire, per ciascun dato, il numero di posizione nella lista ordinata in ordine numerico crescente: il dato più piccolo avrà numero di posizione 1, e via dicendo;
→ quando due o più dati sono uguali, assegnare a ciascuno di essi la media dei numeri di posizione che esse dovrebbero avere; così, per esempio, se i dati dal primo al sesto sono uguali, assegnare come numero di posizione nella lista a tutti il valore 3,5 (media dei numeri da 1 a 6);
→ se i campioni hanno lo stesso numero di dati, calcolare il totale per i numeri di posizione del primo e per i numeri di posizione del secondo campione, e chiamare T il più piccolo di questi due totali;
→ se i due campioni hanno diverso numero di dati chiamare T_1 il totale per il campione che ha il numero minore di dati, diciamo n_1 , e quindi, essendo n_2 il numero di dati del secondo campione, calcolare

$$T_2 = n_1 \cdot (n_1 + n_2 + 1) - T_1$$

→ chiamare T il più piccolo fra i valori T_1 e T_2

→ calcolare la deviata normale standardizzata Z come

$$Z = (| \mu - T | - 0,5) / s$$

essendo

$$\mu = n_1 \cdot (n_1 + n_2 + 1) / 2$$
$$s = \sqrt{(n_2 \cdot \mu / 6)}$$

La statistica Z così calcolata corrisponde a sottoporre al test la mediana delle differenze.

Il valore di p corrispondente alla statistica Z rappresenta la probabilità di osservare per caso una differenza della grandezza di quella effettivamente osservata: se tale probabilità è sufficientemente piccola, si conclude per una significatività della differenza mediana osservata. Questa soluzione è sufficientemente accurata per n > 16.