# Acc-I-4

Misura esistente da rivedere

Definizione: The number of outliers in values is indicating a risk of inaccuracy for data values in a data set. (X=A/B where for X, lower is better)

## Domanda1

Si puo modificare a X = 1-A/B in modo da avere tutte le misure allineate con 0 come valore minimo e peggiore e 1 come valore massimo e migliore?  
  
Risposta: Nello standard Acc-I-4 rappresenta il “Risk of dataset inaccuracy”, per questo ha il valore “migliore” in 0 e il valore “peggiore in 1.

## Domanda2

In questo caso si tratterebbe di una misura conforme che va rinominata e registrata?  
  
Risposta: In alcuni nostri lavori abbiamo usato il suffisso -DevX (con X = A,B,C, etc…) per indicare eventuali variazioni rispetto allo standard, perché non possiamo registrare nuove misure, tuttavia possiamo comunque utilizzare la struttura del naming consigliata (cfr risposta successiva)

## Domanda3

Se si, in base al documento “NAMING e REGISTRAZIONE delle MISURE CONFORMI” deve essere rinominata a Acc-I-4-IT-v ma non saprei quale progressivo mettere al posto della v perche non so se ci sono gia altre misure conformi afferenti al Acc-I-4. (v =numero progressivo per le misure afferenti ad uno stesso valore della stringa precedente (stesso Ccc-Z-Y-AA). Esiste un registro delle misure conformi gia registrate?  
  
Risposta: il registro deve essere ancora approvato, tuttavia possiamo comunque utilizzare la struttura del naming consigliata (cfr risposta successiva). V è un numero progressivo. Parti da 1 e vai avanti se ne fai più di una versione alternativa. Ovviamente è un nostro versionamento interno, perché appunto non esiste ancora un registro.

# Acc-I-3

Nuova misura da includere

Definizione: Data accuracy assurance.  
Ratio of measurement coverage for accurate data.

X=A/B  
A=number of data items measured for accuracy  
B=number of data items for which measurement is required for accuracy

Motivazione: visto che ne abbiamo solo un altra misura di accuracy che calcola gli outlier solo sulle colonne numeriche (utilizzando IQR), mi sembra necessario introdurre questa misura sulla copertura per dare piu contesto alla misurazione attuale dell’accuratezza

Esempio: dataset con 10 colonne dove solo 1 è numerica e non ci sono outlier. Acc-I-4 sarebbe 0 (nella versione originale dove si usa risk, quindi valore basso è migliore) e quindi darebbe l’impressione di una buona accuratezza su questo dataset ma in realtà abbiamo controllato solo una colonna su 10. Copertura sarebbe solo 0.1 e ci dice di “non fidarsi troppo” al valore ottimale ottenuto per Acc-I-4.

## Domanda4

Puo aver senso?

Risposta: Si, può avere senso. Attenzione a cosa si intende per “data item”: lo standard lo definisce come “smallest identifiable unit of data”. Mi sembra meglio associarlo ad un singolo valore di una colonna, piuttosto che alla colonna intera.

# Com-I-1

Misura esistente da rivedere

Definizione: Completeness of data items of a record within a data file

## Domanda5

Nella tesi precedente è stata introdotta la misura derivata Com-I-1 DevA che fa la media su tutti i record del dataset. Pero, nello standard c’è una nota che dice: “This QM can be used to calculate average of completeness for different sets of records”. Considerando questa nota, non sarebbe la misura originale in grado di coprire anche la nostra casistica di fare la media o facendo la media si tratta cmq di una metrica derivata? Se si, va rinominata da Com-I-1 DevA a Com-I-1-IT-v seguendo la nuova naming?

Risposta: Nello standard il “data record” è definito come “set of related data items treated as a unit”. Quindi lo standard prevede sì il calcolo della media, ma all’interno di un singolo record, ottenendo un valore per ogni riga del dataset. Se si vuole una media a livello di dataset, è più opportuno presentare una metrica derivata. Per quanto riguarda il naming, vedi Risposta 3.

# Con-I-5

Nuova misura da includere

Definizione: Data values consistency coverage.

## Domanda6

Dalle 3 misure esistenti sul consistency Con-I-2, Con-I-3, Con-I-4 solo Con-I-3 ha come target le data values. Avrebbe senso anche qua introdurre la copertura come per l’accuracy (domanda4)?

Risposta: Si, ha senso.