

XML Representation of Data Set

A data set is represented as an instance of SpkDataML. This document defines the SpkDataML language.

Table of Contents

SpkDataML DTD.....	1
Example.....	3

SpkDataML DTD

```
<!ELEMENT spkdata (table)*>
<!ATTLIST spkdata version CDATA #REQUIRED>

<!ELEMENT table (description? | weight? | row*)>
<!ATTLIST table columns CDATA #REQUIRED>
<!ATTLIST table rows CDATA #REQUIRED>

<!ELEMENT description (#PCDATA)>

<!ELEMENT weight (#PCDATA)>          --- WARNING! This feature is not yet implemented
<!ATTLIST weight base (sd|fsd|pois|gen) #REQUIRED>
<!ATTLIST weight absolute (yes | no) #REQUIRED>
<!ATTLIST weight x CDATA #REQUIRED>
<!ATTLIST weight y CDATA #IMPLIED>
<!ATTLIST weight z CDATA #IMPLIED>

<!ELEMENT row (value)*>
<!ATTLIST row position CDATA #REQUIRED>

<ELEMENT value (#PCDATA)>
<!ATTLIST value ref CDATA #IMPLIED>
<!ATTLIST value type (numeric|string) #IMPLIED>
```

spkdata

The top element of the SpkDataML document. This can contain multiple tables.

version

The version number of SpkDataML. The initial version is 0.1.

table

A table or matrix of values. The first row in the table is the header row, which means it contains the labels describing the nature of values/data in the columns.

columns

The number of columns in the table, ≥ 0 .

XML Representation of Data Set

rows

The number of rows in the table, ≥ 0 .

description

A short description for the table.

weight

(IMPORATNT! not yet implemented as of 12/13/2005)

A type of standard deviation used to calculate weights on each datum. Select one of the four types as the *base* attribute value.

base = { sd | fsd | pois | gen }

sd

Assign a known standard deviation, x , to each datum.

fsd

Assign a known fractional standard deviation, x , to each datum. $SD = x * | \text{data } i j |$

pois

Assign Poisson statistics. $SD = \sqrt{x * | \text{data } i j |}$

gen

Assign a customized standard deviation using the formula: $SD = \sqrt{x + y * | \text{data } i j | ^z}$

x, y, z

Set x attribute to the value used to calculate SD, FSD and POIS. Set all x, y and z to the values for GEN.

absolute = { yes | no }

yes

The weights are assigned absolutely. They are assumed to be known, and are equal to the reciprocal of the variance of each datum. If " sdi " is the standard deviation of the i -th datum, then the weight associated with the datum is $1/sdi^2$. --- SAAM II User Guide

no

The weights are assigned relatively. They are assumed to be known up to an unknown proportionality constant that is estimated from the data. In this case, if sdi is the standard deviation of the i -th datum in the j -th block of data, the weight associated with the datum is v_j/sdi^2 . The proportionality constant v_j is estimated for each block of data. --- SAAM II User Guide

row

A row in the table.

position

The position of the row in the table, >=1.

value

A value in the row. When the data type is specified as "numeric", only numerical characters (ie. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9) are accepted. For "string" type, only alphabet (ie., a, b, c, ..., A, B, C,...), underscores (ie. _) and spaces are accepted. accepted.

type = { string | numeric }

The data type of the value.

ref

A string associated with the value. This attribute value has no significance; it should be used only as an annotation/note/remark.

Example

An example: Theophylline, 12 subjects¹

Notes

1. ../dataML/data.xml

