XtabML

Survey Table Interchange Format

An XML format for describing the structure and content of survey cross-tabulations.

Version 1.1 **October 2005**

Revision 1.1.000

Authors Geoff Wright & Iain MacKay

Introduction

This document describes the XtabML format for survey cross-tabulation tables.

Background

The aim of the XtabML format is to define a concise means of describing the metadata and contents of typical survey and market research tables.

Summary

The format of the file has been designed to enable software processing routines to be easy to implement (in particular XPath expressions and XSLT processing). To further aid the development process the files are relatively simple to read by eye.

The Definition File

Outline

The definition file is coded in XML syntax according to rules given by the associated XtabML DTD (Document Type Definition). The definition file contents describe three aspects:

a. *the file itself* in terms of version number, date and time of creation etc.

b. *the table report* in terms of the overall settings.

c. *the actual table* or *tables* in terms of the geometry and cell values.

The following shows an outline of the contents of the definition file.

```
<?xml version="1.0"?>

<xtab version="1.0">
       <date>date text</date>
       <origin>origin text</origin>
       . . .
       <control type="DataSource">
              <t>description of data sourcet</t>
       </control>
       . . .
       <statistictype name="ColPct">
              <t>Column %</t>
       </statistictype>
       <table>
              <control type="TableTitle">
                     <t>table title text</t>
              </control>
              . . .
              <edge axis="c">
                     . . .
              </edge>
              . . .
              <statistic type=”ColPct”/>
              . . .
              <data>
                     <r i="1">
                            . . .
                     </r>
                     . . .
              </data>
       </table>
</xtab>
```

Note that the file starts with a declaration that it consists of XML.

The rest of the file is specified in terms of elements such as `<date>` and `<time>`, some of which (such as `<table> ... </table>`) also encapsulate other elements and some of which (such as `<edge axis="c">`) include attributes.

Formatting

1. Reabability

If the definition file may have to be human readable then it is recommended that:

a. The file is organised into lines using CR, LF combinations. However they should be avoided within elements that contain text (e.g. `<t>`) or where their presence could affect how the text is processed.

b. At most one element, or element with associated attributes, appears on one line.

c. Lines are indented with space or tab characters to reflect the structure inherent in the file. An indent is applied after every element that contains other elements.

2. Comments

Comments may be used to annotate contents or to temporarily hide sections of the file from the XML parsing mechanism. These are standard XML comments and start with the conventional XML construct of `<!--` and end with **`-->`**. Comments are optional and can appear any number of times in the definition file.

a. `<!--`*`comment_text`*`-->` can be used anywhere (after the initial <?xml …> declaration) to indicate parts of the definition file that are to be ignored.

b. A *`comment_text`* may include any text except two successive dash characters, `--`.

For example: `<!--Data collected from 12-18th June 2002-->`

3. Unicode

XML uses the Unicode character set. As a character set, Unicode is composed of 17 planes of up to 65,536 characters each. The first plane (plane 00) is called the Basic Multilingual Plane and contains all the characters that most table descriptions would need. In order to represent such a large potential range of characters within a file XML uses "encoded" characters.

The default encoding is UTF-8. In this the ASCII characters 00-7F are represented unchanged, but all other characters are encoded as 2 or more bytes. So if the definition file only contains ASCII 7-bit characters, which includes 0-9, a-z, & A-Z, then the default UTF-8 is suitable.

However, if the variable names or more likely any texts contain 8-bit characters then they must be encoded into UTF-8 or the actual encoding must be specified. The most common encoding is Latin-1 (or more precisely ISO-8859-1) in which case the initial XML declaration should be:-

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
```

More details and resources on Unicode are available from http://www.unicode.org.

Structure

The overall structure of an XtabML definition file is:

Initial information about the definition file (e.g. XtabML version, file creation date and origin)

Description of languages used for alternative texts (optional)

Declaration of potential controls and metadata (e.g. data source, ob title, filter, notes)

Definition of potential statistic types appearing in the report

Actual controls and metadata for the entire tabulation report (e.g. data source, job title)

Description of a table (may be repeated for multiple tables)

- Actual controls and metadata for the table (e.g. base, weight)
- Geometry of the table in terms of edges
  - Each edge defined in terms of groups, elements and summaries
- Statistics (or cell value types) reported in the table
- Table data (i.e. the cell values and status)

The following sections describe the XML elements and attributes for each part of the XtabML definition file in more detail.

Definition File Elements and Attributes

This section describes the syntax and function of each of the XML elements and attributes used to form an XtabML definition file. The elements are shown in the order they are expected in the file.

Initial information

<?xml version="1.0" [encoding="encoding_identifier**"] ?>**

Required. This declares the XML version (which should always be **1.0**) and an optional encoding.

The `encoding` attribute must be specified if the contents of the file are not **UTF-8** (the default). Note that in UTF-8 the ASCII characters 00-7F are represented unchanged. If the file contains other characters then ensure that they are represented as UTF-8, or the correct encoding must be specified.

<xtab version="1.0">

Required. The <`xtab`> element is always required and is used to encapsulate the entire specification document. It contains a mandatory `version` attribute.

The `version` attribute is used to indicate the version of the XtabML format that applies to this description. At present the only valid value is **1.0**. Note that this is a different version number to the XML version declared in the initial processing instruction, although both are currently version 1.0.

<date>*date text***</date>**

Optional. The *`date text`* should represent the date the file was created.

For example: `<date>20 January 2005</date>`

<time>*time text***</time>**

Optional. The *`time text`* should represent the time the file was created.

For example: `<time>18:32</time>`

<origin>*origin text***</origin>**

Optional. The *`origin text`* should describe the originating program.

For example: `<origin>MyProg v3</origin>`

<user>*user text***</user>**

Optional. The *`user text`* should indicate the identifier of the user who created the file.

For example: `<user>A Smith</user>`

Languages used for alternative texts

```
<language lang="language_identifier"
      [base="yes|no"]>language description</language>
```

Optional, repeatable. The text for any title/text `<t>` element and the cell value elements (`<v>`, `<n>`, `<h>` and `<x>`) may contain any number of language-specific text elements. These alternative texts are in addition to the plain or 'base' text that should always be present for these elements. The alternative language texts are specified by one or more `<a>` elements.

For example:
```
<t>Yes<a lang="en-US">Sure</a>
      <a lang="fr">Oui</a>
</t>
```

The `<language>` elements declare the set of potential languages used for these alternative texts.

For example:
```
<language lang="en-US">American</language>
<language lang="fr">French</language>
```

Although there is no restriction on the `language_identifier`, the intended values of the `lang` attribute are the same as the `xml:lang` attribute described in the official W3C XML version 1.0 specification as:-

> "The values of the attribute are language identifiers as defined by IETF (Internet Engineering Task Force) RFC 1766, *Tags for the Identification of Languages* (http://www.ietf.org/rfc/rfc1766.txt) or its successor on the Standards Track.
>
> Note:
>
> IETF RFC 1766 tags are constructed from two-letter language codes as defined by ISO 639 (*Codes for the representation of names of languages*), from two-letter country codes as defined by ISO 3166 (*Codes for the representation of countries and their subdivisions – part 1 (country codes)*), or from language identifiers registered with the Internet Assigned Numbers Authority, *Register of Language Tags*. It is expected that the successor to IETF RFC 1766 will introduce three-letter language codes for languages not presently covered by ISO 639."

The optional `base` attribute can be used to indicate the language for the plain or 'base' text. Note that the default for base is `"no"`, so the attribute is only needed on the actual 'base' language.

For example:
```
<language lang="en" base="yes">English</language>
```

This way of declaring the base language can also be used when there are no alternative languages.

Declaration of controls and metadata

```
<controltype name="control_type_identifier"
      status="primary|secondary">
      <t>control_type_description</t></controltype>
```

Optional, repeatable. There are several types of 'metadata' provided in cross-tabulation reports (e.g. job title, survey name, base, filter, weight), but unfortunately little consensus between different tabulation systems.

The <controltype> elements allow the exporter to define their own vocabulary for controls and metadata. Each <controltype> element defines one type of control or metadata. Providing that each exporting system uses control types consistently, then it will be relatively straightforward to use XSLT to adapt the vocabulary to that of another system (i.e. JobTitle in one system becomes ProjectTitle in another).

The name attribute must be used to give a name that can be used later to identify the purpose of individual control texts.

The status attribute reflects the importance of this control to the creation of a table. For example the table title and filtering is vital and primary, whilst any notes are only descriptive and secondary. When processing an XtabML description from a known source the nature and importance of each control type will already be known. The status attribute is intended primarily where the set of control types are not known, or for the purpose of general XSLT processing.

The <t> element is used to provide a title for labelling the control text in reports and tables. Note that the text for the <t> element may contain any number of language-specific alternative texts.

For example:
```
<controltype name="TableTitle" status="primary">
      <t>Table Title
      <a lang="fr">Titre de Tableau </a></t>
</controltype>
```

Controls and metadata that apply to the entire tabulation report

```
<control [name="control_name"]
      type="control_type_identifier">
      <t>control_description</t></control>
```

Optional, repeatable. The <control> elements declare the control types that apply to the entire set of tables. For example, these could include details of the original survey and the data source, and a copyright notice. Note that a separate set of <control> elements can appear within each table description to specify the characteristics of individual tables.

The name attribute for each control is optional and can be used to give this control a name if this would be meaningful and potentially useful.

The type attribute specifies which control type this is an instance of, and hence the importance and title for the control.

The <t> element is used to provide the actual title or text for the control. Note that the text for the <t> element may contain any number of language-specific alternative texts.

For example:
```
<control type="JobTitle">
      <t>Audi user survey 2004</t>
</control>
```

Declaration of statistic types

```
<statistictype name="statistic_type_identifier"
      <t>statistic_type_description</t></statistictype>
```

Optional, repeatable. There are many types of statistic derived from aggregate data presented in cross-tabulation reports (e.g. unweighted frequency, column percentage). Some of these statistic

types are accumulated from the raw data by a consistent method in all major tabulation systems. Other more complex statistics (e.g. chi-square significance) do not have a completely consistent calculation. XtabML provides a mechanism to name statistic types that should be used consistently across all documents created by a particular accumulation system, and additionally provides a set of standard names (defined in xxx) to cover the commonest statistics.

The `name` attribute is needed to give this statistic type a name which should be consistent across different tables produced by a particular tabulation system. Providing that each exporting system uses statistic types consistently, then it will be relatively straightforward to use XSLT to adapt the vocabulary to that of another system. There is a standard set of statistic type names for common statistic types described in xxx

The `<t>` element provides the title text for reporting this statistic.

For example:
```
<statistictype name="ColPct">
      <t>Column %</t>
</statistictype>
```

Description of each table

Then for each table being described there should be a block comprising:

<table [name="*table_name*"]>

Required, repeatable. Introduces details of an individual table.

The `name` attribute is optional and can be used to give this table a name if this would be meaningful and potentially useful.

<t>*table title***</t>**

Optional. The table title can be specified here as a `<t>` element, or later as a ‘table title’ `<control>` element. The advantage of using the `<t>` element here is that it avoids any need to recognise the ‘table title’ control type from an unknown tabulation system.

Controls and metadata for the table

<control [name="*control_name***"]**
type="*control_type_identifier***">**
<t>*control_description***</t></control>**

Optional, repeatable. These `<control>` elements declare the control types that apply to this table. This could include details of the filters, weights, and notes about this table.

The `name` attribute is optional and can be used to give this control a name if this would be meaningful and potentially useful.

The `type` attribute specifies which control type this is an instance of, and hence the importance and title for the control.

The <t> element is used to provide the actual title or text for the control. Note that the text for the <t> element may contain any number of language-specific alternative texts.

For example:
```
<control type="Filter">
      <t>All over 21 years old</t>
</control>
```

Geometry of the table

<edge [name="*edge_name***"]**
axis="*c|r|p***"**
[level="*hyperplane_level***"]>**

Required, repeatable. The <edge> elements describe the geometry of the table. For example a 2-dimesional table will have one edge for the columns, and one for the rows.

The name attribute is optional and can be used to give this edge definition a name if this would be meaningful and potentially useful. For example, if two or more tables share the same breakdown columns then giving this edge a name may allow a processing program to make use of this information.

The axis attribute specifies which dimension of the table uses this edge. The permitted values are:-

c the columns of the table

r the rows of the table

p the planes of a three or more dimensional table

For example: <edge name="StdBreaks" axis="c">

Note that every table must have at least a "c" edge.

The level attribute is only used when the edge describes the fourth or higher dimension of the table. In this case we use axis="p" level="4" for the fourth dimension etc.

Edge definition (in terms of groups, elements and summaries)

Each <edge> element contains a hierarchy of <group> elements, each of which may contain <summary>, <element> and further <group> elements.

<group [name="*group_name***"]>**

Required. The top level <group> element encloses the details of this edge, and can have an optional name attribute

This top level <group> element will contain some or all of <t>, <element>, <summary> and further <group> elements. The lower level groups are roughly equivalent to variables, and reflect a structural grouping of columns, rows or planes. If a <t> element is present then it should appear first and provides the heading text for this group

For example: <edge axis="c">

```
            <group>
                <summary type="total">
                    <t>Total</t>
                </summary>
                <group name="Region">
                    <t>Region</t>
                    <element><t>London</t></element>
                    <element><t>South East</t></element>
                    ...
                    <summary type="net">
                        <t>All England</t>
                    </summary>
                    <element><t>Borders</t></element>
                    <element><t>Glasgow</t></element>
                    ...
                    <summary type="net">
                        <t>All Scotland</t>
                    </summary>
                    ...
                </group>
                <group name="Age">
                    ...
                </group>
            </group>
        </edge>
```

<element [name="*element_name***"]>**
`<`**t**`>`*element description*`</`**t**`>`**</element>**

Optional, repeatable. Each `<element>` element describes a row, column or plane that contains ordinary data. As such an `<element>` is roughly equivalent to the 'category' of a variable, but is more general because it can also be something like a quantity.

The `name` attribute is optional and can be used to give this element definition a name if this would be meaningful and potentially useful.

The `<t>` element provides the heading text for this column, row or plane.

For example:

```
<group name="Region">
    <t>Region</t>
    <element><t>London</t></element>
    <element><t>South East</t></element>
    ...
```

<summary [name="*summary_name***"]**
type="*summary_type***"]>**
`<`**t**`>`*summary description*`</`**t**`>`**</summary>**

Optional, repeatable. Each `<summary>` element describes a special row, column or plane that summarises the members of the group in some way.

The `name` attribute is optional and can be used to give this summary definition a name if this would be meaningful and potentially useful.

The optional `type` attribute can be used to specify the type of summary. For example this might include `type="total"` and `type="net"`. XtabML: standardises only one value for this attribute, i.e. `type="xs:base"` to be used on summary elements that are the base for percentaging within their group.

The `<t>` element provides the text for this column, row or plane.

For example:
```
<group name="Region">
      <t>Region</t>
      <element><t>London</t></element>
      <element><t>South East</t></element>
      ...
      <summary type="net"><t>All England</t></summary>
      ...
```

Statistics (cell values) within the table

<statistic type="*statistic_type_name***"**
[datatype="*integer|decimal|string|percentage***"]**
[scale="*scale factor***"]**
[display="*yes|no***"]/>**

Required, repeatable. The `<statistic>` elements declare the types and order of statistics (or cell values) that appear within this table.

The name used in the mandatory **`type`** attribute should be one of the statistic types defined in the **`xtab`** element.

The presence of a **`datatype`** attribute signals that the values of this statistic in the data element of the table will be in canonical format.

A **`scale`** attribute gives the value of a scale factor by which to multiply the canonical value to obtain the value of the statistic – the scale factor is optional and defaults to 1.0.

Occasionally, especially for integer values, the desired formatted appearance of the statistic will be identical to the canonical format. In this situation the attribute **`display="yes"`** may be used to signal that the values of the statistic may be interpreted as canonical or as formatted.

An exporting application may output the same statistic twice if required, once as formatted and once as canonical, though in most applications it is anticipated that XtabML documents will contain only formatted or only canonical values.

If none of the optional attributes appear than the statistic will be rendered in the data section as a formatted value.

Table data

<data>

Required. The <data> element specifies the values and status for each cell of the table. These cells are presented in a nested format with columns (<c> elements) nested within rows (<r> elements) within planes (<p> elements).

For example:

```
<data>
    <r i="1">
        <c>
        ...
        </c>
    </r>
    <r i="2">
        <c>
        ...
        </c>
    </r>
    ...
</data>
```

The nested representation accords with the natural ordering of the cells in a table, and provides efficient access to cell values. Within each row, there is a <c> element for each statistic defined for the table. The row elements do not correspond to lines in a printed table but to logical rows that may contain multiple statistics.

There is a one-to-one correspondence between <element> and <summary> elements within the row edge definition. and <r> elements within the <data> element. Similarly, there is a one-to-one correspondence between sub-elements of each <c> element and <element> and <summary> elements within the c edge definition, and similarly for planes and higher dimensions. <group> elements within edge definitions do not affect the <data> element – they are only present to document the structural relationship of <element> and <summary> elements within the edge.

<p [i="*index*">]
[l="*level*">]>

Optional, repeatable. The <p> elements contain the values for one plane or hyperplane. If it is a simple plane then it will contain row values, but if it is a fourth or higher dimension plane then it will contain the values for the next lower level plane.

The i attribute is optional and can be used to indicate the index of this plane. As the format of the data section is fairly dense, the index can be useful when inspecting or debugging the data values.

The l attribute specifies the dimension level if this is a fourth or higher dimension plane.

<r [i="*index*">]>

Optional, repeatable. The <r> element contains the values for one row. It will contain the statistics values for the columns.

The `i` attribute is optional and can be used to indicate the index of this row. As the format of the data section is fairly dense, the index can be useful when inspecting or debugging the data values.

`<c>``...`**`</c>`**

Required, repeatable. The `<c>` element contains the column values for each statistic. So within each row there will be one `<c>..</c>` block for each statistic. Within this block are the actual values for that statistic.

Statistics values and status

The statistic values and status are indicated by four element types:-

- **`<v>`** *`cell value`***`</v>`** A normal printed statistic value.
- **`<h>`** *`cell value`***`</h>`** A statistic value that is not intended for printing (e.g. unweighted values in a weighted table, or statistics for a suppressed row).
- **`<x>`** *`cell value`***`</x>`**, or **`<x/>`** A statistic that is not applicable and the cell should not appear (e.g. the mean for a categorical row).
- **`<n>`** *`cell value`***`</n>`**, or **`<n/>`** A statistic that is not applicable but the geometry of the table means that the cell should appear.

The `<x/>` and `<n/>` representations should be used unless the exporting system has a preferred method of representing inapplicable cell statistics (e.g. 'n/a'). For most conventional tables the `<v>..</v>` and `<x/>` cells will be most common.

Normally a statistic will have the same status for every column, but this is not required. In fact it is likely that any `<n/>` cells will be intermixed with `<v>` cells.

The content of these elements represents either the formatted value of the cell (i.e. the value that would appear within a printed report), or the canonical value, depending on the d`atatype` attribute of the applicable `statistic` element, if present. The text for these elements when they show canonical values may contain any number of language-specific alternative texts where different formatting rules apply to different locations (e.g. the use of '.' or ',' as a decimal place separator).

In order to save space within the data section only statistic value elements up to the last distinct element need be included within each `<c>..</c>`; i.e. if the row ends with a run of identical elements this may be reduced to a single element. This compression is particularly useful when the entire statistic consists of `<x/>` elements. Note that this compression is optional for exporting systems.

For example:

```
<r i="1">
    <c>
        <v>54</v><v>12</v><v>9</v><v>16</v>
        <v>17</v><v>12</v><v>24</v><v>16</v>
    </c>
    <c><h>100.00</h></c>
    <c><x/></c>
</r>
<r i="2">
    ..
```

Representation of canonical values within v and h elements

On the principle of reusing existing standards we looked at using part of the W3C standard “XML Schema Part 2: Datatypes” (http://www.w3.org/TR/xmlschema-2/) as the basis for defining the lexical representation of the data types. This standard provides (amongst others) decimal, integer and string data types that fit very closely the proposed XtabML data types.

decimal

[Definition:] decimal represents arbitrary precision decimal numbers. The ‘value space’ of decimal is the set of the values i × 10^-n, where i and n are integers such that n >= 0. The ‘order-relation’ on decimal is: x < y iff y - x is positive.

[Lexical representation:] decimal has a lexical representation consisting of a finite-length sequence of decimal digits (#x30-#x39) separated by a period as a decimal indicator. … An optional leading sign is allowed. If the sign is omitted, "+" is assumed. Leading and trailing zeroes are optional. If the fractional part is zero, the period and following zero(es) can be omitted. For example: -1.23, 12678967.543233, +100000.00, 210.

[Note:] W3C dos not have special values like NaN and INF for decimal (these are used with type float). If the value can not be represented as a decimal then the <v> element should be empty. Note that values that are incalculable for structural reasons should be represented as empty <x/> elements.

integer

[Definition:] integer is ‘derived’ from decimal by fixing the value of ‘fractionDigits’ to be 0. This results in the standard mathematical concept of the integer numbers. The ‘value space’ of integer is the infinite set {...,-2,-1,0,1,2,...}. The ‘base type’ of integer is decimal.

[Lexical representation:] integer has a lexical representation consisting of a finite-length sequence of decimal digits (#x30-#x39) with an optional leading sign. If the sign is omitted, "+" is assumed. For example: -1, 0, 12678967543233, +100000.

[Note:] W3C dos not have special values like NaN and INF for decimal (these are used with type float). If the value can not be represented as a decimal then the <v> element should be empty. Note that values that are incalculable for structural reasons should be represented as empty <x/> elements.

string

[Definition:] The string datatype represents character strings in XML. The ‘value space’of string is the set of finite-length sequences of characters (as defined in [XML 1.0 (Second Edition)]) that match the Char production from [XML 1.0 (Second Edition)]. A character is an atomic unit of communication; it is not further specified except to note that every character has a corresponding Universal Character Set code point, which is an integer.

percentage

Percentage does not have a standard W3C definition, but as this data type only represents different formatting of a decimal value it will share the same definition and lexical representation.

[Note:] A percentage is in effect a ratio that is displayed with 100 understood as the denominator. Whilst most percentages are proportions which range from 0..1 (e.g. column percent), some reflect ratios that may be negative or exceed 1. The use of the term percentage here does not imply any 0..1 restriction on the canonical values.

Example

Example XtabML Definition File

The example defines a single table:

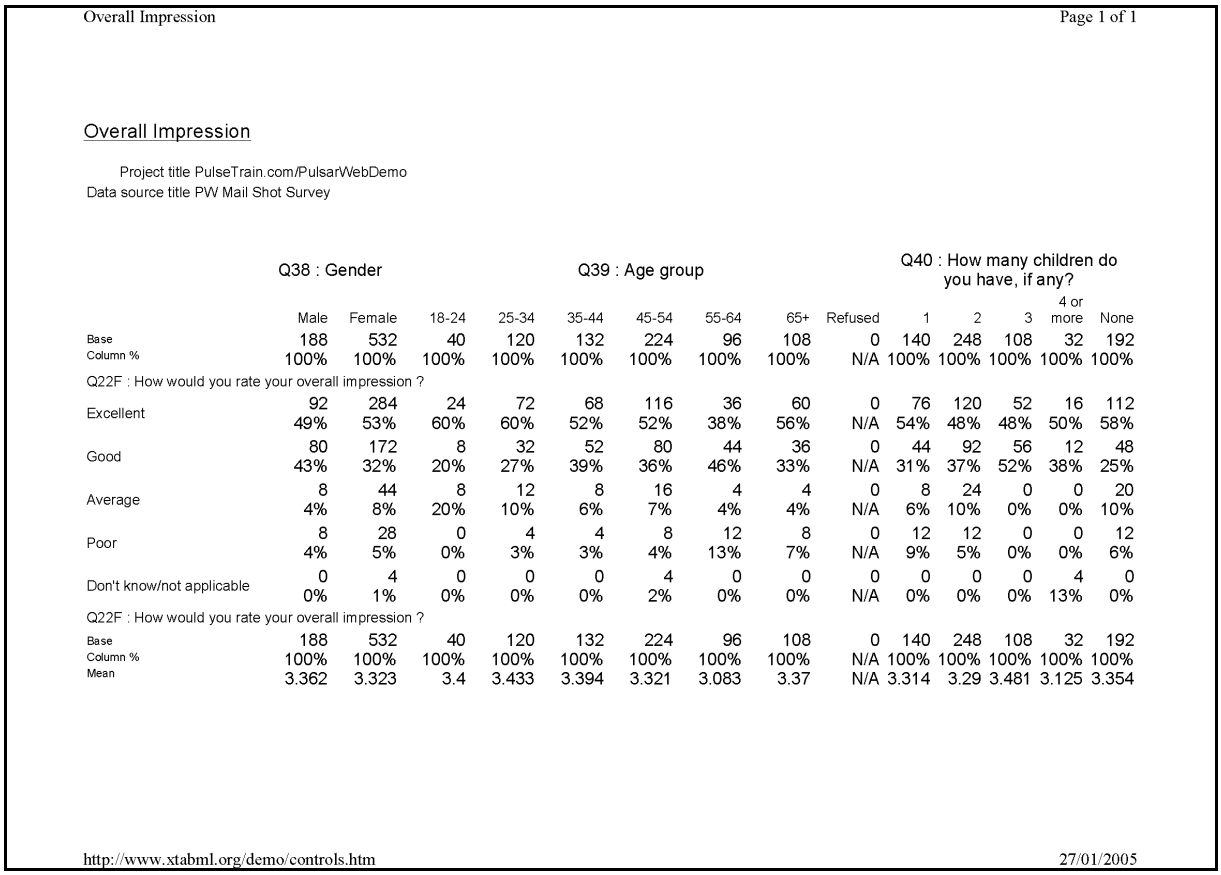
```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE xtab PUBLIC "-//XtabML//DTD Cross-tabulation reports v1.1//EN"
                      "http://www.xtabml.org/dtd/xtabml_v11.dtd">
<xtab version="1.1"
  xmlns:xt="http://www.XtabML.org/2005/xtab"
  xmlns="http://www.XtabML.org/2005/xtab">

    <date>2005/01/10</date>
    <time>14:45:00</time>
    <origin>A tabulation system</origin>

    <controltype name='pt' status='primary'><t>Project title</t></controltype>
    <controltype name='st' status='primary'><t>Data source title</t></controltype>
    <controltype name='gf' status='primary'><t>Global filter</t></controltype>
    <controltype name='pf' status='primary'><t>Pre-filter</t></controltype>
    <controltype name='f' status='primary'><t>Filter</t></controltype>
    <controltype name='dw' status='primary'><t>Data source weight</t></controltype>
    <controltype name='uw' status='primary'><t>User weight</t></controltype>
    <controltype name='notes' status='secondary'><t>Notes</t></controltype>

    <!--XtabML standard statistics-->
    <statistictype name="xs:rt"><t>Unweighted sample</t></statistictype>
    <statistictype name="xs:wt"><t>Weighted sample</t></statistictype>
    <statistictype name="xs:wsqt"><t>Sum of squared weights</t></statistictype>
    <statistictype name="xt:t"><t>Total</t></statistictype>
    <statistictype name="xs:qt"><t>Total volume</t></statistictype>
    <statistictype name="xs:qsqt"><t>Total squared volume</t></statistictype>
    <statistictype name="xs:ess"><t>Effective sample size</t></statistictype>
    <statistictype name="xs:mean"><t>Mean</t></statistictype>
    <statistictype name="xs:variance"><t>Variance</t></statistictype>
    <statistictype name="xs:sd"><t>S.D.</t></statistictype>
    <statistictype name="xs:se"><t>S.E.</t></statistictype>
    <statistictype name="xs:sew"><t>S.E.</t></statistictype>
    <statistictype name="xs:cp"><t>Column %</t></statistictype>
    <statistictype name="xs:rp"><t>Row %</t></statistictype>
    <statistictype name="xs:pp"><t>Table %</t></statistictype>
    <statistictype name="xs:lcp"><t>Row group %</t></statistictype>
    <statistictype name="xs:lrp"><t>Column group %</t></statistictype>
    <statistictype name="xs:qcp"><t>Column % (volume)</t></statistictype>
    <statistictype name="xs:qrp"><t>Row % (volume)</t></statistictype>
    <statistictype name="xs:qpp"><t>Table % (volume)</t></statistictype>
    <statistictype name="xs:lqcp"><t>Row group % (volume)</t></statistictype>
    <statistictype name="xs:lrcp"><t>Column group % (volume)</t></statistictype>
    <statistictype name="xs:pop"><t>Population</t></statistictype>
    <!--End of XtabML standard statistics-->

    <control type="pt"><t>PulseTrain.com/PulsarWebDemo</t></control>
    <control type="st"><t>PW Mail Shot Survey</t></control>

    <table>
        <t>Overall Impression</t>
        <edge axis="r">
            <group>
                <summary type="xs:base"><t>Total</t></summary>
                <group><t>Q22F : How would you rate your overall impression ?</t>
                    <element score="4"><t>Excellent</t></element>
                    <element score="3"><t>Good</t></element>
                    <element score="2"><t>Average</t></element>
                    <element score="1"><t>Poor</t></element>
```

```
                <element><t>Don't know/not applicable</t></element>
            </group>
            <group><t>Q22F : How would you rate your overall impression ?</t>
                <summary/>
            </group>
        </group>
    </edge>
    <edge axis="c">
        <group>
            <group><t>Q38 : Gender</t>
                <element><t>Male</t></element>
                <element><t>Female</t></element>
            </group>
            <group><t>Q39 : Age group</t>
                <element><t>18-24</t></element>
                <element><t>25-34</t></element>
                <element><t>35-44</t></element>
                <element><t>45-54</t></element>
                <element><t>55-64</t></element>
                <element><t>65+</t></element>
                <element><t>Refused</t></element>
            </group>
            <group><t>Q40 : How many children do you have, if any?</t>
                <element><t>1</t></element>
                <element><t>2</t></element>
                <element><t>3</t></element>
                <element><t>4 or more</t></element>
                <element><t>None</t></element>
            </group>
        </group>
    </edge>

    <statistic type="xs:t" datatype="integer" display="yes"/>
    <statistic type="xs:cp"/>
    <statistic type="xs:mean"/>

    <data>
        <r i='1'>
            <c>
                <v>188</v><v>532</v><v>40</v><v>120</v><v>132</v>
                <v>224</v><v>96</v><v>108</v><v>0</v><v>140</v>
                <v>248</v><v>108</v><v>32</v><v>192</v>
            </c>
            <c>
                <v>100%</v><v>100%</v><v>100%</v><v>100%</v><v>100%</v>
                <v>100%</v><v>100%</v><v>100%</v><v>N/A</v><v>100%</v>
            </c>
            <c><x/></c>
        </r>
        <r i='2'>
            <c>
                <v>92</v><v>284</v><v>24</v><v>72</v><v>68</v>
                <v>116</v><v>36</v><v>60</v><v>0</v><v>76</v>
                <v>120</v><v>52</v><v>16</v><v>112</v>
            </c>
            <c>
                <v>49%</v><v>53%</v><v>60%</v><v>60%</v><v>52%</v>
                <v>52%</v><v>38%</v><v>56%</v><v>N/A</v><v>54%</v>
                <v>48%</v><v>48%</v><v>50%</v><v>58%</v>
            </c>
            <c><x/></c>
        </r>
        <r i='3'>
            <c>
                <v>80</v><v>172</v><v>8</v><v>32</v><v>52</v>
                <v>80</v><v>44</v><v>36</v><v>0</v><v>44</v>
                <v>92</v><v>56</v><v>12</v><v>48</v>
            </c>
```

```
                <c>
                    <v>43%</v><v>32%</v><v>20%</v><v>27%</v><v>39%</v>
                    <v>36%</v><v>46%</v><v>33%</v><v>N/A</v><v>31%</v>
                    <v>37%</v><v>52%</v><v>38%</v><v>25%</v>
                </c>
                <c><x/></c>
            </r>
            <r i='4'>
                <c>
                    <v>8</v><v>44</v><v>8</v><v>12</v><v>8</v>
                    <v>16</v><v>4</v><v>4</v><v>0</v><v>8</v>
                    <v>24</v><v>0</v><v>0</v><v>20</v>
                </c>
                <c>
                    <v>4%</v><v>8%</v><v>20%</v><v>10%</v><v>6%</v>
                    <v>7%</v><v>4%</v><v>4%</v><v>N/A</v><v>6%</v>
                    <v>10%</v><v>0%</v><v>0%</v><v>10%</v>
                </c>
                <c><x/></c>
            </r>
            <r i='5'>
                <c>
                    <v>8</v><v>28</v><v>0</v><v>4</v><v>4</v>
                    <v>8</v><v>12</v><v>8</v><v>0</v><v>12</v>
                    <v>12</v><v>0</v><v>0</v><v>12</v>
                </c>
                <c>
                    <v>4%</v><v>5%</v><v>0%</v><v>3%</v><v>3%</v>
                    <v>4%</v><v>13%</v><v>7%</v><v>N/A</v><v>9%</v>
                    <v>5%</v><v>0%</v><v>0%</v><v>6%</v>
                </c>
                <c><x/></c>
            </r>
            <r i='6'>
                <c>
                   <v>0</v><v>4</v><v>0</v><v>0</v><v>0</v>
                   <v>4</v><v>0</v><v>0</v><v>0</v><v>0</v>
                   <v>0</v><v>0</v><v>4</v><v>0</v>
                </c>
                <c>
                    <v>0%</v><v>1%</v><v>0%</v><v>0%</v><v>0%</v>
                    <v>2%</v><v>0%</v><v>0%</v><v>N/A</v><v>0%</v>
                    <v>0%</v><v>0%</v><v>13%</v><v>0%</v>
                </c>
        <c><x/></c>
            </r>
            <r i='7'>
                <c>
                    <v>188</v><v>532</v><v>40</v><v>120</v><v>132</v>
                    <v>224</v><v>96</v><v>108</v><v>0</v><v>140</v>
                    <v>248</v><v>108</v><v>32</v><v>192</v>
                </c>
                <c>
                    <v>100%</v><v>100%</v><v>100%</v><v>100%</v><v>100%</v>
                    <v>100%</v><v>100%</v><v>100%</v><v>N/A</v><v>100%</v>
                </c>
                <c>
                    <v>3.362</v><v>3.323</v><v>3.4</v><v>3.433</v><v>3.394</v>
                    <v>3.321</v><v>3.083</v><v>3.37</v><v>N/A</v><v>3.314</v>
                    <v>3.29</v><v>3.481</v><v>3.125</v><v>3.354</v>
                </c>
            </r>
         </data>
    </table>
</xtab>
```

Interpretation

The previous data corresponds to a table that could be printed as follows:



Overall Impression — Page 1 of 1

Overall Impression

Project title PulseTrain.com/PulsarWebDemo
Data source title PW Mail Shot Survey

	Q38 : Gender		Q39 : Age group						Q40 : How many children do you have, if any?					
	Male	Female	18-24	25-34	35-44	45-54	55-64	65+	Refused	1	2	3	4 or more	None
Base	188	532	40	120	132	224	96	108	0	140	248	108	32	192
Column %	100%	100%	100%	100%	100%	100%	100%	100%	N/A	100%	100%	100%	100%	100%
Q22F : How would you rate your overall impression ?														
Excellent	92	284	24	72	68	116	36	60	0	76	120	52	16	112
	49%	53%	60%	60%	52%	52%	38%	56%	N/A	54%	48%	48%	50%	58%
Good	80	172	8	32	52	80	44	36	0	44	92	56	12	48
	43%	32%	20%	27%	39%	36%	46%	33%	N/A	31%	37%	52%	38%	25%
Average	8	44	8	12	8	16	4	4	0	8	24	0	0	20
	4%	8%	20%	10%	6%	7%	4%	4%	N/A	6%	10%	0%	0%	10%
Poor	8	28	0	4	4	8	12	8	0	12	12	0	0	12
	4%	5%	0%	3%	3%	4%	13%	7%	N/A	9%	5%	0%	0%	6%
Don't know/not applicable	0	4	0	0	0	4	0	0	0	0	0	0	4	0
	0%	1%	0%	0%	0%	2%	0%	0%	N/A	0%	0%	0%	13%	0%
Q22F : How would you rate your overall impression ?														
Base	188	532	40	120	132	224	96	108	0	140	248	108	32	192
Column %	100%	100%	100%	100%	100%	100%	100%	100%	N/A	100%	100%	100%	100%	100%
Mean	3.362	3.323	3.4	3.433	3.394	3.321	3.083	3.37	N/A	3.314	3.29	3.481	3.125	3.354

http://www.xtabml.org/demo/controls.htm 27/01/2005

The XtabML DTD

Summary

The XtabML 1.0 DTD is given below. As with all XML code, this document is required if the syntax of an XtabML XML description file is to be verified as 'valid' rather than simply being considered 'well formed'. Note that this DTD is available online from http://www.xtabml.org/dtd/xtabml_v10.dtd.

```
<!-- XtabML.dtd                                                    -->
<!-- Standalone XML version 1.1                                    -->
<!-- October 2005                                                    -
->
<!-- ============================================================  -->
<!-- An XML definition for storing cross-tabulation reports from   -->
<!-- a tabulation system.                                          -->
<!-- ============================================================  -->
<!-- This DTD has been produced by Computable Functions Ltd        -->
<!-- for Pulse Train Ltd.                                          -->
<!-- ============================================================  -->
<!-- Public identifier:                                            -->
<!--   -//XtabML//DTD Cross-tabulation reports v1.1//EN            -->
<!-- Public URL:                                                   -->
<!--   http://www.xtabml.org/dtd/xtabml_v11.dtd                    -->
<!-- Public namespace:                                             -->
<!--   xmlns:xt="http://www.xtabml.org/2005/xtab"
-->
<!-- ============================================================  -->
<!--                                                               -->
<!-- Version history:                                              -->
<!--                                                               -->
<!--   1.0   Apr 2005  Original version                            -->
<!--   1.1   Oct 2005  Canonical values                            -->
<!--                                                               -->
<!-- ============================================================  -->
<!--                                                               -->
<!--          BEGINNING OF ACTUAL DOCUMENT TYPE DEFINITION         -->
<!--                                                               -->

<!-- parameter entities                                           -->
<!ENTITY % datatype "decimal |
                     integer |
                     percent |
                     string" >

<!-- multi-language alternative texts                             -->
<!ELEMENT a ( #PCDATA ) >
<!ATTLIST a
          lang NMTOKEN #REQUIRED>
<!ENTITY % texts "( #PCDATA | a )*">


<!-- TOP LEVEL                                                    -->
<!-- =========                                                    -->
<!--                                                              -->

<!ELEMENT xtab ( date?, time?, origin?, user?,
                 language*, controltype*, statistictype+,
                 control*, table+ ) >
<!ATTLIST xtab
          version NMTOKEN #REQUIRED>

    <!ELEMENT date ( #PCDATA ) >

    <!ELEMENT time ( #PCDATA ) >
```

```
    <!ELEMENT origin ( #PCDATA ) >

    <!ELEMENT user ( #PCDATA ) >

<!-- Define the languages used within multi-lingual texts          -->

    <!ELEMENT language ( #PCDATA ) >
    <!ATTLIST language
              lang NMTOKEN #REQUIRED
              base ( yes | no ) "no" >


<!-- CONTROL TYPES DESCRIPTION                                     -->
<!-- ========================                                      -->
<!--                                                               -->
<!-- Controls describe items like the survey title, data source,   -->
<!-- filters, weights, and other notes about the table.            -->
<!-- The controltype name declares the controls available to this  -->
<!-- set of tables and their textual description.                  -->

    <!ELEMENT controltype (t ) >
    <!ATTLIST controltype
              name NMTOKEN #REQUIRED
              status ( primary | secondary ) #REQUIRED >

<!-- The t element is used universally for element texts/titles    -->
<!-- which can always be multi-lingual.                            -->

        <!ELEMENT t %texts; >


<!-- STATISTIC TYPES DESCRIPTION                                   -->
<!-- ===========================                                   -->
<!--                                                               -->
<!-- Statistics are the values calculated for each cell of the     -->
<!-- table (e.g. col%, weighted count, mean).                      -->
<!-- The statistictype name declares the staistics available to    -->
<!-- this set of tables, and their textual description.            -->

    <!ELEMENT statistictype ( t ) >
    <!ATTLIST statistictype
              name NMTOKEN #REQUIRED >

        <!-- ELEMENT t already defined -->


<!-- CONTROLS DESCRIPTION                                          -->
<!-- ====================                                          -->
<!--                                                               -->
<!-- Define the external controls used by this table (i.e. survey  -->
<!-- title, data source etc). Each control will have a type (from  -->
<!-- controltype), a name, and a descriptive text.                 -->

    <!ELEMENT control ( t ) >
    <!ATTLIST control
              name NMTOKEN #IMPLIED
              type NMTOKEN #REQUIRED >

        <!-- ELEMENT t already defined -->


<!-- TABLE DESCRIPTION                                             -->
<!-- =================                                             -->
<!--                                                               -->

    <!ELEMENT table ( t?, control*, edge+, statistic+, data? ) >
    <!ATTLIST table
```

```
              name NMTOKEN #IMPLIED >

        <!-- ELEMENT t already defined -->


<!-- TABLE CONTROLS                                                    -->
<!-- ==============                                                    -->
<!--                                                                   -->
<!-- Define the internal controls used within this table (i.e.        -->
<!-- filter, weight etc). Each control will have a type (from         -->
<!-- controltype), a name, and a descriptive text.                     -->

        <!-- ELEMENT control already defined -->



<!-- TABLE EDGES                                                       -->
<!-- ===========                                                       -->
<!--                                                                   -->
<!-- Define the edges used to create this table, and the axes         -->
<!-- used for the data element.                                        -->

        <!ELEMENT edge ( t?, group ) >
        <!ATTLIST edge
                  name NMTOKEN #IMPLIED
                  axis ( c | r | p ) #REQUIRED
                  level NMTOKEN #IMPLIED >

            <!-- ELEMENT t already defined -->

            <!ELEMENT group ( t | group | element | summary )* >
            <!ATTLIST group
                      name NMTOKEN #IMPLIED >

                <!-- ELEMENT t already defined -->

                <!ELEMENT element ( t ) >
                <!ATTLIST element
                          name NMTOKEN #IMPLIED
                          score CDATA #IMPLIED >

                    <!-- ELEMENT t already defined -->

                <!ELEMENT summary ( t? ) >
                <!ATTLIST summary
                          name NMTOKEN #IMPLIED
                          type NMTOKEN #IMPLIED >

                    <!-- ELEMENT t already defined -->


<!-- TABLE STATISTICS                                                  -->
<!-- ================                                                  -->
<!--                                                                   -->
<!-- Define the statistics present in this table (i.e. count, col%)   -->
<!-- their order and representation.                                   -->

        <!ELEMENT statistic EMPTY >
        <!ATTLIST statistic
                  type NMTOKEN #REQUIRED
                  datatype ( %datatype; ) #IMPLIED
                  scale CDATA #IMPLIED
                  display ( yes | no ) #IMPLIED >


<!-- TABLE DATA                                                        -->
<!-- ==========                                                        -->
<!--                                                                   -->
```

```
<!-- Element and attribute names for the table data have been       -->
<!-- deliberately chosen to be single characters in order to save   -->
<!-- space.                                                         -->
<!--     table geometry:-                                           -->
<!--         c - column                                             -->
<!--         r - row                                                -->
<!--         p - plane (optional level used if more than one plane) -->
<!--     statistic values:-                                         -->
<!--         v - valid value                                        -->
<!--         n - no value                                           -->
<!--         h - value is hidden                                    -->
<!--         x - no value and hidden                                -->

        <!ELEMENT data ( p+ | r+ | c+ ) >

            <!ELEMENT p ( p+ | r+ ) >          <!-- plane -->
            <!ATTLIST p
                      i NMTOKEN #IMPLIED
                      l NMTOKEN #IMPLIED >       <!-- index & level -->

                <!ELEMENT r ( c+ ) >               <!-- row -->
                <!ATTLIST r
                          i NMTOKEN #IMPLIED >           <!-- index -->

                    <!ELEMENT c ( ( v | n | h | x )+ ) >   <!-- col -->

                        <!ELEMENT v ( #PCDATA ) >  <!-- valid value -->

                        <!ELEMENT n EMPTY >     <!-- not applicable -->

                        <!ELEMENT h ( #PCDATA ) > <!-- hidden value -->

                        <!ELEMENT x EMPTY >       <!-- hidden & N/A -->
```

XtabML standard statistics

Optional

An XtabML document is not obligated to use any of the standard statistic names, even for statistics whose calculation conforms to one of the standard names. However, if a statistic is given one of the standard names, it must have exactly the meaning described here.

A statistic may be present in a table without a value being provided in all cells of the table; many statistic types such as populations or weighted sums of squares are typically available or calculated only in the margins of the table.

Prefix

An XtabML document may declare amongst its statistic types any of the standard ones described here. The names of all standard statistics begn with the prefix "**xs**:". Any statistic name declared having this prefix must be one of the standard names.

Notation

Statistic values always appear in a cell of the table, and each cell has an associated position in each of the edges providing the dimensions of the table.

The edge definitions contain a mixture of **element** and **summary** components. Each component can be identified by its position in the edge, and has an associated condition that qualifies respondents to be accumulated at that position. The definition of conditions is outside the scope of XtabML. The position of an **element** or **summary** component is its position in the sequence of **element** and **summary** children of the **edge** XML element, ignoring the **group** XML elements recording the hierarchy of edge components..

All XtabML statistic values appear in cells. Each cell has a position in each of the edges of the table. In the definitions following, r stands for a row position, and c for a column position.

So, for example **xs:t** (r, c) represents the value of standard statistic type **xs:t** at row r and column c. None of the standard statistic values requires values from more than one plane, so it is not necessary to include a plane or higher dimension subscript.

The notation **xs:t**(*,c) represents the value of statistic **xs:t** calculated using the qualification for column c and no row qualification, i.e. the value of the statistic in the base column (which may or may not be a column actually stored in the XtabML document).

Finally the notation **xs:t**(r*,c) represents the value of statistic **xs:t** calculated in the base row for the group containing row r, e.g. an "all answering" total. Again, this may or may not be a row that is actually stored: XtabML does not mandate that each group has a summary).

Recommended titles

XtabML allows the exporter to choose a title for each statistic type used. This reflects the reality that there is more variation in the titles for statistics (even in the same natural language) than there is variation in the methods of calculation. For convenience, XtabML provides a recommended English language title for each statistic.

For volume statistics in particular, applications are likely to choose their own project-specific titles.

Data types

Standard statistic types are intended for applications that will do calculations with the statistic values, or store them in a database or some other function that requires interpretation of the value. So standard statistics must be output in canonical format, and therefore the **statistic** elements that reference them must have an XtabML **datatype** attribute. The **datatype** attribute value for each standard value is shown in the summary at 0.

Basic statistics

In this section some statistics are defined in terms of other statistics. The definitions are purely to express the meaning of the statistics. It is not mandatory to include a statistic type in the XtabML document just because it is used in the definition of another statistic that is intended for output. Indeed several of the statistics defined such as sums of squares are rarely published. They are included in the standard because of their role as intermediate values in calculating the other statistics that are published. However, all statistic types that are included should be mutually consistent according to these definitions.

xs:rt (unweighted total)

xs:rt(r, c) is the unweighted or 'raw' count of respondents qualifying for both row r and column c.

If the table is weighted in row r and column c, then only respondents having a weight should be counted.

Additionally, if there is a quantity applying to row r and column c, then only those respondents with a defined value for the quantity should be counted.

These rules are necessary to ensure that effective sample size for weights, and dispersion statistics for quantities, will be correctly calculated.

Note that consequently rows with different quantities applying but with the same row condition may have different unweighted totals.

xs:t (publication total)

xs:t(r, c) is either the sum of respondent weights for all respondents qualifying for row r, column c if there is a weight associated with row r and column c or otherwise the value of **xs:rt** (r, c), multiplied by any factor projecting the weights to population values (if the surveys totals are to be published as population estimates).

If there is a quantity associated with row r column c then only respondents possessing a value of the quantity will be included.

In a typical formatted table the totals in detail rows will be **xs:t** statistics, while in the base row **xs:rt** statistics are presented as well as **xs:t**.

xs:qt (quantity total)

xs:qt(r, c) is the sum of quantity values for all respondents qualifying for row r and column c. If there is a weight associated with row r and column c, then **xs:qt** is the sum of the product of the respondent weight and the quantity for each respondent.

Note that in tables derived from scored questions the quantity value is calculated from a frequency distribution by assigning scores to each category of the distribution. This mechanism is outside the scope of XtabML; XtabML simply records the outcome of the calculation.

xs:qsqt (sum of squared quantities)

xs:qsqt(r, c) is the sum of squared quantity values for all respondents qualfying for row r and column c. If there is a weight associated with row r anc column c, then **xs:qsqt** is the sum of the product of the respondent weight and the squared quantity for each respondent.

xs:wsqt (sum of squared weights)

xs:wsqt (r, c) is the sum of squared weights for respondents qualifying for row r and column c.

xs:wt (weighted total)

xs:wt(r, c) is the sum of unprojected respondent weights for all respondents qualifying for row r, column c if there is a weight associated with row r and column c; otherwise **xs:wt** is equal to **xs:rt**.

If there is a quantity associated with row r column c then only respondents possessing a value of the quantity will be included. By unprojected we mean that the weights used have an average of one and so this statistic is effectively a weighted sample count.

xs:wt will not normally be published, but may be required for the calculation of the weighted standard error statistic which requires an unprojected weighted total. Occasionally when projected figures are published, the reports include a row of 'weighted sample counts' along with the unweighted totals, all other totals in the tables being projected to the universe of the survey. **xs:wt** in this situation is the 'weighted sample count' statistic.

Derived statistics

Derived statistics are calculated from the basic accumulated statistics.

In the calculations below, if for any cell the calculation leads to an undefined value, e.g. division by zero, then the value for the statistic is treated as incalculable and exported as an empty **v** or **h** element. Any statistic then derived from this statistic will also be incalculable and will also be exported as an empty element.

The definitions of the basic statistics were chosen so that calculations of the derived statistics do not differ if weights are present or absent.

xs:ess (effective sample size)

xs:ess (r, c) is the effective sample size in row r and column c. It is calculated as: **xs:ess** (r, c) **= xs:t****2/ **xs:wsqt**(r, c), if there is weighting applicable to row r, column c, or as **xs:rt** otherwise.

This definition permits **xs:ess** to serve as the 'N' in later statistical calculations whether or not weighting is used.

Dispersion statistics

xs:mean (quantity mean)

xs:mean (r,c) is defined as **xs:qt**(r, c) / **xs:t**(r, c).

xs:variance (quantity variance)

xs:variance (r, c) is calculated as: (**xs:qsqt** (r, c) - **xs:qt**(r,c)**2/ **xs:t**) / (**xs:wt** - 1)

xs:sd (quantity standard deviation)

xs:sd (r, c) is defined as sqrt(**xs:variance**(r, c))

xs:se (quantity standard error)

xs:se (r, c) is defined as **xs:sd** (r, c) / sqrt(**xs:ess** (r, c))

xs:sew (quantity standard error, weighted)

Some tabulation systems do not have the sum of squared weights available in every context where a standard error is required. These systems typically use the (unprojected) weighted total when calculating the standard error. XtabML allows for this with a special value type **xs:sew**, calculated as: **xs:sew** (r, c) = **xs:sd**(r, c)/sqrt(**xs:wt**(r,c)) xs:sew values are not best practice and **xs:se** values should be published instead wherever possible.

Percentages

There are potentially very many percentage statistic types, but only a few arise in practice. There are two main sources of variation of percentage calculation, namely the statistic type being percentaged, and where to find the base for the percentage.

The statistic being percentaged is normally the unweighted total, weighted total or quantity total. However, if tables are weighted it is uncommon to show unweighted percentages, or to calculate an unweighted quantity total for quantity percentages

Identifying the percentage base is more complex. Each cell is at a position in the hierarchy of groups in each edge of the table. So there are bases available in each enclosing group in each edge, potentially a large number of choices. The XtabML standard statistics include percentages using the **xt:t** (publication total) statistic and **xt:qt** (quantity total) statistic only, based only on five possibilities to select the base. This still creates ten percentage statistic types. Note that by using the publication total XtabML standard percentages are always weighted if the table is weighted, and unweighted otherwise.

xs:cp (column percentage)

xs:cp (r,c) is defined as: 100 * **xs:t**(r, c**)/xs:t**(*, c)

xs:rp (row percentage)

xs:rp (r, c) is defined as: **xs:t**(r, c) **/ xs:t**(r, *)

xs:pp (page or plane percentage)

xs:pp (r, c) is defined as: 100 * **xs:t**(r, c) **/ xs:t**(*, *)

xs:lcp (local column percentage)

xs:lcp (r,c) is defined as: 100 * **xs:t**(r, c**)/xs:t**(r*, c)

'local' percentages typically arise when 'all answering the question' is the intended base.

xs:lrp (local row percentage)

xs:lrp (r, c) is defined as: **xs:t**(r, c) **/ xs:t**(r, c*)

xs:qcp (quantity column percentage)

xs:cp (r,c) is defined as: 100 * **xs:qt**(r, c**)/xs:qt**(*, c)

xs:qrp (quantity row percentage)

xs:rp (r, c) is defined as: **xs:qt**(r, c) **/ xs:qt**(r, *)

xs:qpp (page or plane quantity percentage)

xs:pp (r, c) is defined as: 100 * **xs:qt**(r, c) **/ xs:qt**(*, *)

xs:lqcp (local quantity column percentage)

xs:lcp (r,c) is defined as: 100 * **xs:qt**(r, c**)/xs:qt**(r*, c)

'local' percentages typically arise when 'all answering the question' is the intended base.

xs:lqrp (local quantity row percentage)

xs:lrp (r, c) is defined as: **xs:qt**(r, c) **/ xs:qt**(r, c*)

External statistics

There are some statistic types that do not arise from accumulation and come into the table from external data. Presently there is only one of these defined as a standard statistic.

xs:pop (Population)

xs:pop (r, c) is the population (or universe) of respondents qualfying for row r and column c. Typically populations are only available in the margins of tables, so an exported table will include for population only **xs:pop** (*, c) or **xs:pop** (r, *) or even **xs:pop**(*, *). If weighting is in use and the sample has been 'projected' to the universe then the xs:t values may be identical or very close to the **xs:pop** values. The exporting application determines whether the weighted totals are projected or unprojected. Typically, if population figures are available, then weighted totals in XtabML documents will be projected.

Exclusions

Medians and other 'quantile' statistics derived from frequency distributions are not included in the standard statistics, because of diversity in the methods used to estimate them; i.e. interpolate/don't interpolate etcetera. Significance test results are not standardised for similar reasons.

Relationship of score values to standard statistics

It is not mandatory to include the score values that gave rise to statistic values in an XtabML document. They are convenient for consuming applications. However, if score values are included, then their relationship to the associated quantity statistics is as follows:

- the score values and their associated quantity statistics will be in the same edge **group** (whether row, column or hyper/plane).
- the quantity statistics will be reported in a **summary** component of the **group**
- only **element** components of the **group** that have scores will be included in the base for the statistics

These rules enable scores to be used in any dimension of the table. If scores are used in more than one dimension (rows and columns, say), the quantity values in the intersection of the summary statistic row and summary statistic column are undefined in XtabML and should be exported as ‘not available’.

Summary of standard statistics

Name	Standard title	Data type	Notes
xs:rt	Unweighted sample	**integer**	
xs:wt	Weighted sample	**decimal**	unprojected weighted total
xs:wsqt	Sum of squared weights	**decimal**	
xt:t	Total	**decimal**	includes any projection factor if weighted
xs:qt	Total volume	**decimal**	.
xs:qsqt	Total squared volume	**decimal**	
xs:ess	Effective sample size	**decimal**	sample efficiency = xs:ess / xs:rt
xs:mean	Mean	**decimal**	
xs:variance	Variance	**decimal**	
xs:sd	S.D.	**decimal**	
xs:se	S.E.	**decimal**	
xs:sew	S.E.	**decimal**	
xs:cp	Column %	**percent**	
xs:rp	Row %	**percent**	
xs:pp	Table %	**percent**	
xs:lcp	Row group %	**percent**	
xs:lrp	Column group %	**percent**	
xs:qcp	Column % (volume)	**percent**	
xs:qrp	Row % (volume)	**percent**	
xs:qpp	Table % (volume)	**percent**	
xs:lqcp	Row group % (volume)	**percent**	
xs:lrcp	Column group % (volume)	**percent**	

xs:pop	Population	**integer**	