Description

unicode encoding list and unicode encoding alias list encodings that are available in Stata. See help encodings for advice on choosing an encoding and a list of the most common encodings. unicode encoding list provides a list of all encodings and their aliases or those that meet specified criteria. unicode encoding alias provides a list of alternative names that may be used to refer to a specific encoding.

unicode encoding set sets an encoding to be used with the unicode translate command; see [D] **unicode translate** for documentation for unicode encoding set.

Syntax

List encodings

unicode <u>enc</u>oding list [*pattern*]

List all aliases of an encoding

unicode <u>enc</u>oding alias *name*

Set an encoding for use with unicode translate

unicode <u>enc</u>oding set *name*

pattern is one of the following: *, _all, *name*, *name, or name*. Specifying nothing, _all, or * lists all results. Specifying *name* lists all results containing *name*. Specifying *name lists all results ending with *name*. Specifying name* lists all results starting with *name*.

Remarks and examples

Encoding is the method by which text is stored in a computer. It maps a character to a nonnegative integer, called a code point, then maps that integer to a single byte or a sequence of bytes. Common encodings are ASCII (for which there are many variants), UTF-8, and UTF-16. Stata uses UTF-8 encoding for storing text and UTF-16 to encode the GUI on Microsoft Windows and macOS. For more information about encodings, see [U] **12.4.2.3 Encodings**.

The most common reason you will need to specify an encoding is when converting a dataset, do-file, ado-file, or some other file used with Stata 13 or earlier (which was not Unicode aware) for use with modern Stata. See [D] **unicode translate** for help with this, and see help encodings for advice on choosing an encoding and a list of common encodings.

Some commands and functions require that you specify one or more encodings. Often you will need to use only common encodings. However, you may not know how to specify these to Stata. For example, suppose that we are using `unicode translate` to convert a do-file from Stata 13 that contains extended ASCII characters for use in modern Stata. If we are working on a Windows machine, the most likely encoding is Windows-1252. If we want to check that this is how it should be specified as we use `unicode translate`, we can type

        . unicode encoding list Windows-1252

Stata returns all encodings for which the encoding name or an alias exactly matches `Windows-1252`. Capitalization does not matter.

If we wanted to search for all encodings and aliases that have `windows` anywhere in their name, we could type

        . unicode encoding list *windows*

and see a long list of matches.

If we are told that a text file is encoded with `ibm-913_P100-2000` and we want to see by what other names that encoding is known (perhaps because we just do not want to type out such a long string when using Stata's functions that need an encoding), we can use

        . unicode encoding alias ibm-913_P100-2000

and we find that there are many synonyms, including some that are much easier to type.

You may not know the exact encoding that you need and wish to browse the full list of available encodings. To do this, you can just type `unicode encoding list` without specifying a pattern.

Also see

`help encodings`

[D] **unicode** — Unicode utilities

[D] **unicode translate** — Translate files to Unicode

[U] **12.4.2 Handling Unicode strings**

[U] **12.4.2.3 Encodings**