

ustrunescape() — Convert escaped hex sequences to Unicode strings

Description Syntax Remarks and examples Conformability
 Diagnostics Also see

Description

`ustrunescape(s)` returns a string of Unicode characters corresponding to the escaped sequences in *s*.

`ustrtohex(s [, n])` returns a string of escaped hex digit Unicode characters, up to 200, specified in *s*. If *n* is specified and is larger than one, the result starts at the *n*th Unicode character of *s*; otherwise, it starts at the first Unicode character.

When *s* is not a scalar, these functions return element-by-element results.

Syntax

string matrix `ustrunescape(string matrix s)`

string matrix `ustrtohex(string matrix s [, real scalar n])`

Remarks and examples

stata.com

The following escape sequences are recognized by `ustrunescape()`:

4 hex digit form	<code>\uhhhh</code>
8 hex digit form	<code>\Uhhhhhhhh</code>
1–2 hex digit form	<code>\xhh</code>
1–3 octal digit form	<code>\ooo</code>

where *h* is in `[0–9A–Fa–f]` and *o* is in `[0–7]`. The standard ANSI C escape sequences `\a`, `\b`, `\t`, `\n`, `\v`, `\f`, `\r`, `\e`, `\"`, `\'`, `\?`, and `\\` are recognized as well. The function returns an empty string if an escape sequence is badly formed. Note that the 8 hex digit form `\Uhhhhhhhh` starts with a capital “U”.

`ustrtohex()` converts each Unicode character in string *s* to an escaped hex string. Unicode code points in the range `\u0000–\uffff` are converted to 4 hex digit form `\uhhhh`. Unicode code points greater than `\uffff` are converted to 8 hex digit form `\Uhhhhhhhh`.

`ustrtohex()` converts an invalid UTF-8 sequence to the Unicode replacement character `\ufffd`.

The null terminator `char(0)` is a valid Unicode character and its escaped form is `\u0000`.

Conformability

`ustrunescape(s), ustrtohex(s [, n])`:

<i>s</i> :	$r \times c$
<i>n</i> :	1×1
<i>result</i> :	$r \times c$

Diagnostics

None.

Also see

[M-5] [ustrfix\(\)](#) — Replace invalid UTF-8 sequences in Unicode string

[M-5] [ustrto\(\)](#) — Convert a Unicode string to or from a string in a specified encoding

[M-4] [string](#) — String manipulation functions

[U] [12.4.2 Handling Unicode strings](#)