

udsubstr() — Extract Unicode substring based on display columns

Description
Diagnostics

Syntax
Also see

Remarks and examples

Conformability

Description

`udsubstr(s, n1, n2)` returns the Unicode substring of *s*, starting at Unicode character *n1*, for *n2* display columns. If *n2* = . (*missing*), the remaining portion of the Unicode string is returned. If *n2* display columns from Unicode character *n1* is in the middle of a Unicode character, the substring stops at the previous Unicode character.

When arguments are not scalar, `udsubstr()` returns element-by-element results.

Syntax

string matrix `udsubstr(string matrix s, real matrix n1, real matrix n2)`

Remarks and examples

stata.com

n1 < 0 is interpreted as distance from the end of the Unicode string; *n1* = -1 means starting at the last Unicode character.

An invalid UTF-8 sequence is replaced with a Unicode replacement character `\ufffd`. Null terminator `char(0)` in a binary string is a valid UTF-8 character and will be counted and treated as such.

Use `ustr()` to extract a substring based on Unicode characters. Use `substr()` to extract a substring based on bytes.

Conformability

`udsubstr(s, n1, n2):`

<i>s</i> :	$r \times c$
<i>n1</i> :	$r \times c$ or 1×1
<i>n2</i> :	$r \times c$ or 1×1
<i>result</i> :	$r \times c$

Diagnostics

`udsubstr(s, n1, n2)` returns an empty string if an error occurs.

Also see

[M-5] **subinstr()** — Substitute text

[M-5] **substr()** — Extract substring

[M-5] **_substr()** — Substitute into string

[M-5] **usubinstr()** — Replace Unicode substring

[M-5] **usubstr()** — Extract Unicode substring

[M-5] **_usubstr()** — Substitute into Unicode string

[M-4] **string** — String manipulation functions

[U] **12.4.2 Handling Unicode strings**