

ustr substr() — Extract Unicode substring

[Description](#) [Syntax](#) [Remarks and examples](#) [Conformability](#)
[Diagnostics](#) [Also see](#)

Description

`ustr substr(s, n1, n2)` returns the Unicode substring of *s*, starting at Unicode character *n1*, for a length of *n2*. If *n1* < 0, *n1* is interpreted as the distance from the last Unicode character of *s*; if *n2* = . (*missing*), the remaining portion of the Unicode string is returned.

`ustr left(s, n)` returns the first *n* Unicode characters of the Unicode string *s*.

`ustr right(s, n)` returns the last *n* Unicode characters of the Unicode string *s*.

When arguments are not scalar, the functions return element-by-element results.

Syntax

string matrix `ustr substr(string matrix s, real matrix n1, real matrix n2)`

string matrix `ustr left(string matrix s, real matrix n)`

string matrix `ustr right(string matrix s, real matrix n)`

Remarks and examples

[stata.com](#)

$n \leq 0$ is interpreted as the distance from the end of the Unicode string; $n = -1$ means the distance 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 `udsubstr()` to extract a substring based on display columns. Use `substr()` to extract a substring based on bytes.

Conformability

`ustr substr(s, b, l)`:

<i>s</i> :	$r \times c$
<i>b</i> :	$r \times c$ or 1×1
<i>l</i> :	$r \times c$ or 1×1
<i>result</i> :	$r \times c$

Diagnostics

`ustrsubstr(s, b, l)`, `ustrleft(s, b, l)`, and `ustrright(s, b, l)` return 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] **ustrsubstr()** — Replace Unicode substring

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

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

[U] **12.4.2 Handling Unicode strings**

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow and NetCourseNow are trademarks of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.



For suggested citations, see the FAQ on [citing Stata documentation](#).