

unicode — Unicode utilities

[Description](#) [Remarks and examples](#) [Also see](#)

[[Suggestion](#): Read [[U](#)] [12.4.2 Handling Unicode strings](#) first.]

Description

The `unicode` command provides utilities to help you work with Unicode strings in your data. If you have only plain ASCII characters in your data (a–z, A–Z, 0–9, and typical punctuation characters), you can stop reading now. Otherwise, continue with *Remarks and examples* below.

Remarks and examples

stata.com

We recommend that you start with some overview documentation. First, you should read [[U](#)] [12.4.2 Handling Unicode strings](#), which will explain the difference between ASCII and Unicode and provide detailed advice on working with Unicode strings in Stata. In that section, you will learn about locales, encodings, sorting, and Unicode-specific string functions. For a general overview of Unicode-specific advice, see `help unicode advice`.

Second, if you have datasets, do-files, ado-files, or other files that you used with Stata 13 or earlier and those files contain characters other than plain ASCII such as accented characters, Chinese, Japanese, or Korean (CJK) characters, Cyrillic characters, and the like, you should read [[D](#)] [unicode translate](#).

`unicode` provides the following utilities:

[D] unicode translate	Translate files to Unicode
[D] unicode encoding	Unicode encoding utilities
[D] unicode locale	Unicode locale utilities
[D] unicode collator	Language-specific Unicode collators
[D] unicode convertfile	Low-level file conversion between encodings

You may also find `help encodings` useful if you need to choose an encoding when converting a string from extended ASCII to Unicode.

Also see

- [[D](#)] [unicode collator](#) — Language-specific Unicode collators
- [[D](#)] [unicode convertfile](#) — Low-level file conversion between encodings
- [[D](#)] [unicode encoding](#) — Unicode encoding utilities
- [[D](#)] [unicode locale](#) — Unicode locale utilities
- [[D](#)] [unicode translate](#) — Translate files to Unicode
- [[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](#).