### Description

tokens(s) returns the contents of s, split into words.

tokens(s, parsechars) returns the contents of s split into tokens based on parsechars.

tokens(s) is equivalent to tokens(s, " ").

If you need more advanced parsing, see [M-5] tokenget().

### Syntax

\[
\begin{align*}
\text{string rowvector} & \quad \text{tokens} \left( \text{string scalar } s \right) \\
\text{string rowvector} & \quad \text{tokens} \left( \text{string scalar } s, \text{ string scalar } \text{ parsechars} \right)
\end{align*}
\]

### Remarks and examples

tokens() is commonly used to split a string containing a sequence of variable names into a row vector, each element of which contains one variable name:

\[\text{tokens}("mpg weight displacement") = ("mpg", "weight", "displacement")\]

Some Stata interface functions require that variable names be specified in this form. This is required, for instance, by st_varindex(); see [M-5] st_varindex(). If you had a string scalar vars containing one or more variable names, you could obtain their variable indices by coding

\[\text{indices} = \text{st_varindex(tokens(vars))}\]

### Conformability

\[
\begin{align*}
tokens(s, \text{ parsechars}) \\
s: & \quad 1 \times 1 \\
\text{parsechars}: & \quad 1 \times 1 \quad \text{(optional)} \\
\text{result}: & \quad 1 \times w, \quad w = \text{number of words (tokens) in } s
\end{align*}
\]

### Diagnostics

If s contains "", tokens() returns \(J(1,0,"")\).

If s contains double-quoted or compound-double-quoted material, the quotes are stripped and that material is returned as one token. For example,

\[\text{tokens('"this "is an" example"')} = ("this", "is an", "example")\]
If \( s \) contains quoted material and the quotes do not match, results are as if the appropriate number of close quotes were added to the end of \( s \). For example,

\[
\text{tokens('"this "is an example"')} = ("this", "is an example")
\]

**Also see**

[M-5] invtokens( ) — Concatenate string rowvector into string scalar

[M-5] tokenget( ) — Advanced parsing

[M-5] ustrword( ) — Obtain Unicode word from Unicode string

[M-4] String — String manipulation functions

[P] gettoken — Low-level parsing

[P] tokenize — Divide strings into tokens