%xusplit - Text Variable Splitting Utility

Overview

The `%xusplit` macro splits long text variables into multiple shorter sub-variables while preserving word boundaries. It's particularly useful for handling long text fields that need to be split across multiple columns while maintaining readability and preventing word truncation.

Version Information

- **Version**: 1.0
- **Last Updated**: 15JUL2022
- **Author(s)**: Atorus Research

Dependencies

- SAS version: SAS 9.4 V9
- No macro dependencies

Parameters

Parameter	Required	Default	Description
inds	Yes	-	Name of the input dataset
invar	Yes	-	Name of the input variable to split
outds	No	&inds	Name of the output dataset
prefix	No	&invar	Prefix for the output variable names
len	No	200	Length of the output variables
debug	No	N	Whether to retain temporary datasets (Y/N)

Return Values/Output

Creates a dataset containing:

- Original dataset variables
- Split text variables named:
    - [prefix] (original variable)
    - [prefix]1 to [prefix]N (additional segments)
- Each split variable limited to specified length
- Words kept intact (not truncated)
- Empty split variables for short text

Processing Details

1. Input Validation:

- Verifies required parameters
- Checks for non-empty values

2. Text Processing:

- Creates observation order tracking
- Removes extra spaces with COMPBL
- Splits text at word boundaries
- Ensures splits don't exceed length limit
- Preserves word integrity

3. Variable Creation:

- Transposes split segments into variables
- Determines required number of variables
- Applies consistent length to all segments
- Merges results with original dataset

Examples

Basic Usage

```
%xusplit(co, coval);
```

Custom Output and Prefix

```
%xusplit(
    inds=dv1,
    invar=term,
    outds=dv,
    prefix=dvterm
);
```

Specified Length with Debug

```
%xusplit(
    inds=trt_final,
    invar=col3,
    len=100,
    debug=Y
);
```

Common Issues and Solutions

Issue	Solution

Issue	Solution
Missing words	Increase len parameter
Too many variables	Increase len to reduce splits
Variable name conflicts	Use unique prefix

Notes and Limitations

- Maximum length for split variables is 32767
- Original variable length is set to len parameter
- Empty strings are preserved in output
- Word boundaries are preserved
- Extra spaces are removed
- Debug mode retains intermediate datasets
- All split variables have same length
- Missing values handled appropriately

See Also

- %xusupp: Supplemental qualifier creation
- %xumrgcs: Dataset merging utility
- %xusave: Dataset saving utility

Change Log

Version	Date	Author	Changes
1.0	15JUL2022	Atorus Research	Initial version