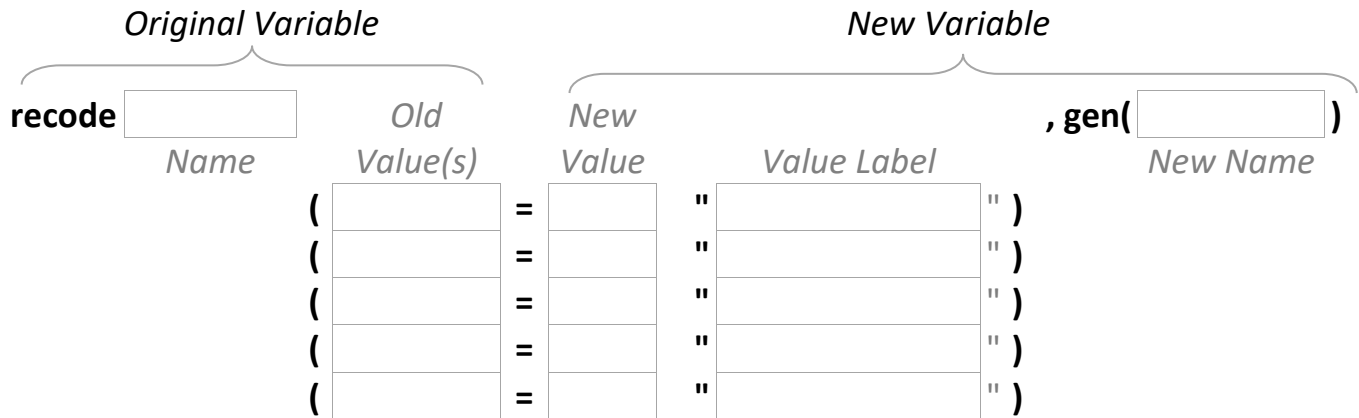


## Worksheet for Recoding in Stata

- ✓ Remember to include the “gen” at the end in order to make a new variable.
- ✓ If you think you might forget, first clone the variable, then use the cloned variable to recode:

**clonevar**  =   
*New Variable Name* *Old Variable Name*

### 1 Categorical Variable → 1 Categorical Variable

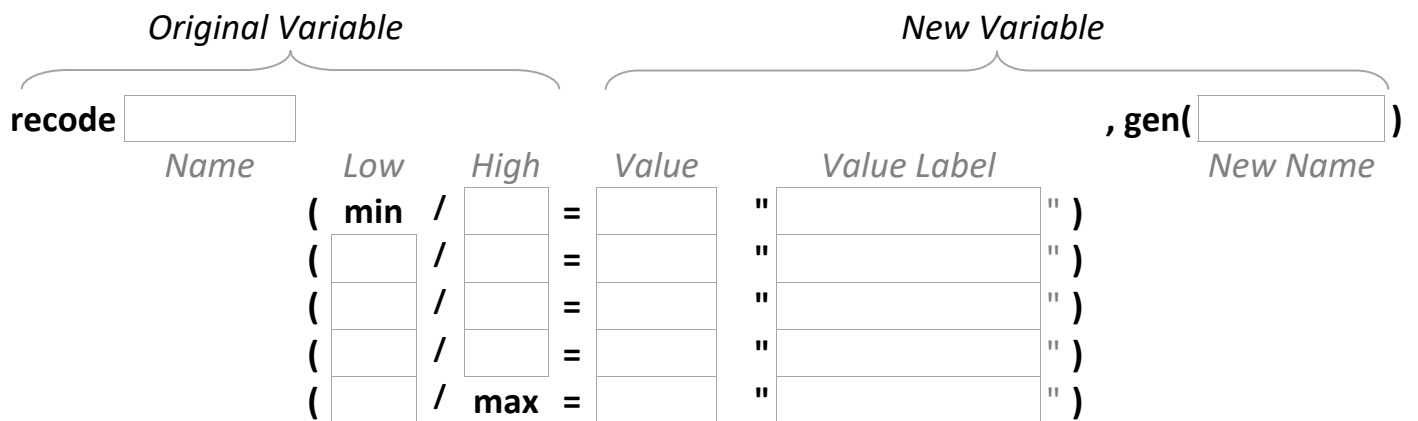


Examples:

recode *score* ( 1 2 = 1 "Low" ) ( 3 = 2 "Med" ) ( 4 5 = 3 "High" ), gen( *score\_in3* )

recode *region* ( 1 = 1 "Northeast" ) ( 2 3 4 = 0 "Not NE" ), gen( *northeast* )

### 1 Continuous Variable → 1 Categorical Variable



Examples:

recode *mpg* (min/18=1 "Low") (19/23=2 "Med") (24/max=3 "High"), gen(*mpg\_in3*)

recode *demstate* ( min/49.9 = 0 "Low" ) ( 50/max = 1 "High" ), gen( *demstate\_in2* )

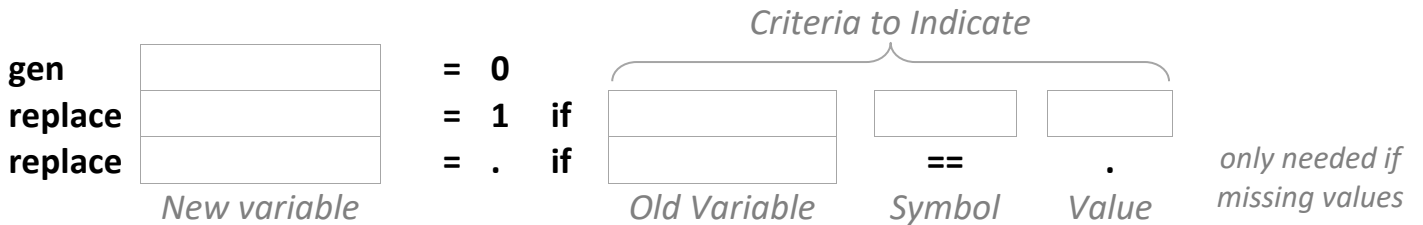
# Worksheet for Dummy Coding in Stata

Symbols: < > == != <= >=

## Original

Pros: simple & straightforward

Cons: much longer with missing values



Examples:

gen *northeast* = 0

replace *northeast* = 1 if *region* == 1

replace *northeast* = . if *region* == .

gen *midwest* = 0

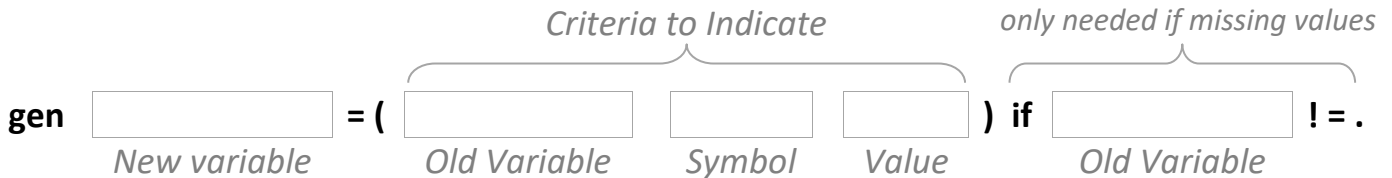
replace *midwest* = 1 if *region* == 2

replace *midwest* = . if *region* == .

## Alternative 1

Pros: short & flexible

Cons: unclear, awkward with missing values



Examples:

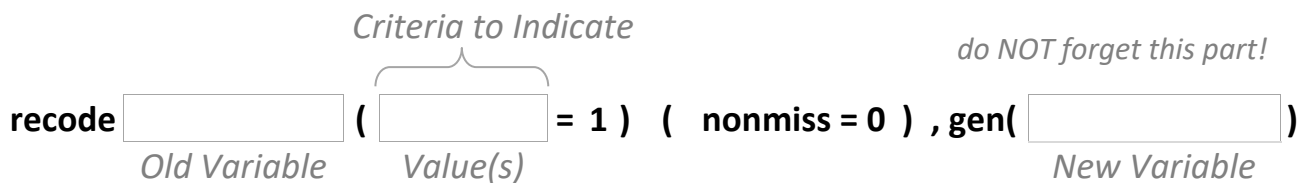
gen *northeast* = ( *region* == 1 ) if *region* != .

gen *midwest* = ( *region* == 2 ) if *region* != .

## Alternative 2

Pros: clear & compact, can label

Cons: rigid criteria, forgetting gen is easy & bad



Examples:

recode *region* ( 1 = 1 ) ( nonmiss = 0 ) , gen( *northeast* )

recode *region* ( 2 = 1 ) ( nonmiss = 0 ) , gen( *midwest* )

or

recode *region* ( 1 = 1 "Northeast" ) ( nonmiss = 0 "Not NE" ) , gen( *northeast* )

recode *region* ( 2 = 1 "Midwest" ) ( nonmiss = 0 "Not MW" ) , gen( *midwest* )