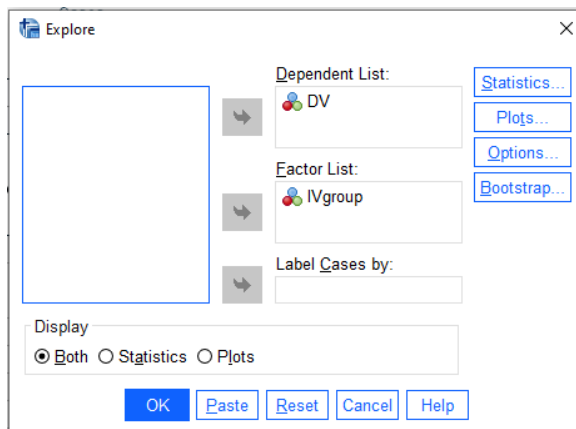


## Trimming & Winsorizing for 2BG ANOVA Designs

When performing data cleaning for ANOVA designs, we have to perform the outlier analysis and any trimming or Winsorizing operations ***separately for each group!!***

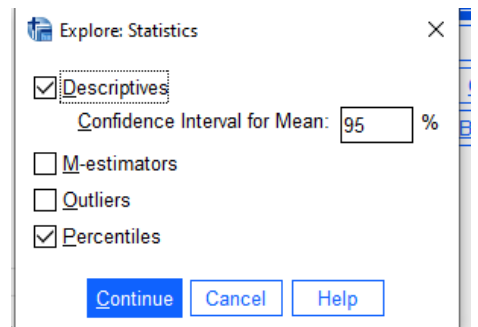
Here is our starting data...

	IVgroup	DV	var	var
1	1.00	9.00		
2	1.00	11.00		
3	1.00	7.00		
4	1.00	12.00		
5	1.00	20.00		
6	1.00	13.00		
7	2.00	20.00		
8	2.00	29.00		
9	2.00	28.00		
10	2.00	26.00		
11	2.00	34.00		
12	2.00	27.00		



We use the “Factor List” window to get a separate analysis for each condition of the variable “IVgroup”

We check to get both descriptives and percentiles.



And the univariate data for each group...

Oneway

Descriptives

DV			
	N	Mean	Std. Deviation
1.00	6	12.0000	4.47214
2.00	6	27.3333	4.54606
Total	12	19.6667	9.08879

The percentiles for the groups are...

Percentiles

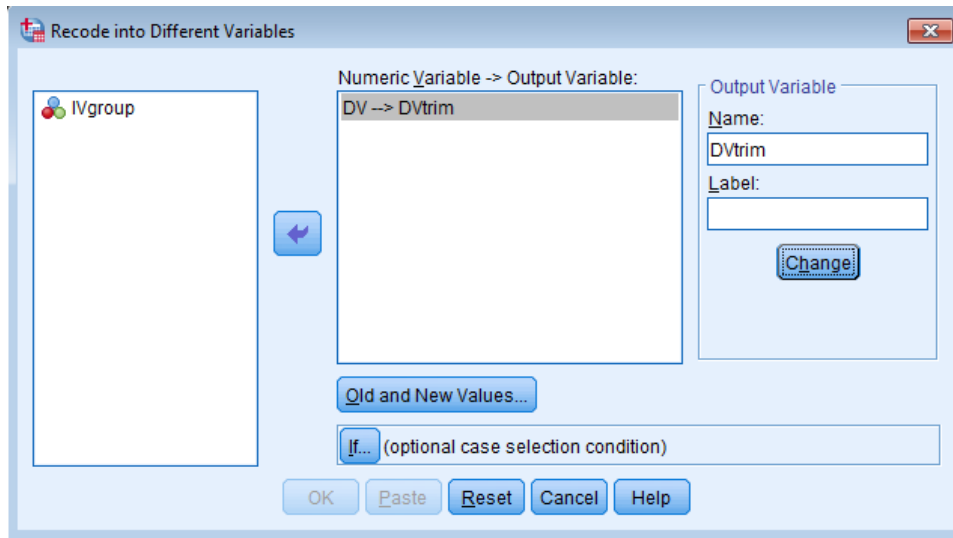
		Percentiles							
		IVgroup	5	10	25	50	75	90	95
Weighted Average (Definition 1)	DV	1.00	7.0000	7.0000	8.5000	11.5000	14.7500	.	.
		2.00	20.0000	20.0000	24.5000	27.5000	30.2500	.	.
Tukey's Hinges	DV	1.00			9.0000	11.5000	13.0000		
		2.00			26.0000	27.5000	29.0000		

Giving outlier boundaries of

For Group 1	For Group 2
<div>Outlier Boundry Calculator</div> <div>Q1 or Lower Hinge =&gt; 9 13 &lt;= Q3 or Lower Hinge</div> <div>Lower Boundry = 3 19 = Upper Boundry</div>	<div>Outlier Boundry Calculator</div> <div>Q1 or Lower Hinge =&gt; 26 29 &lt;= Q3 or Lower Hinge</div> <div>Lower Boundry = 21.5 33.5 = Upper Boundry</div>

## Trimming

The potentially confusing part is that we have to perform the trimming separately for each group, because the two groups have different outlier boundaries. We will do two separate “runs” using different groups, but using the same variable.



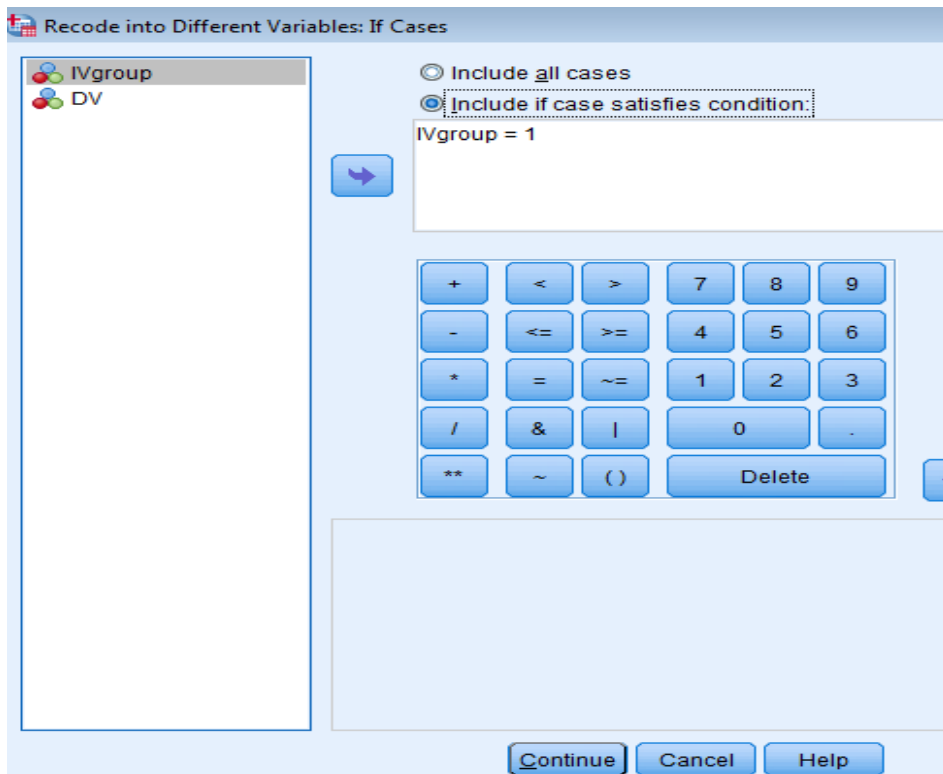
Move the Numeric variable into the textbox.

Specify the name of the Output variable.

Usually we would next click on “Old and New Values”...

But instead, we are going to click on “If...”

## We'll do Group = 1 first...



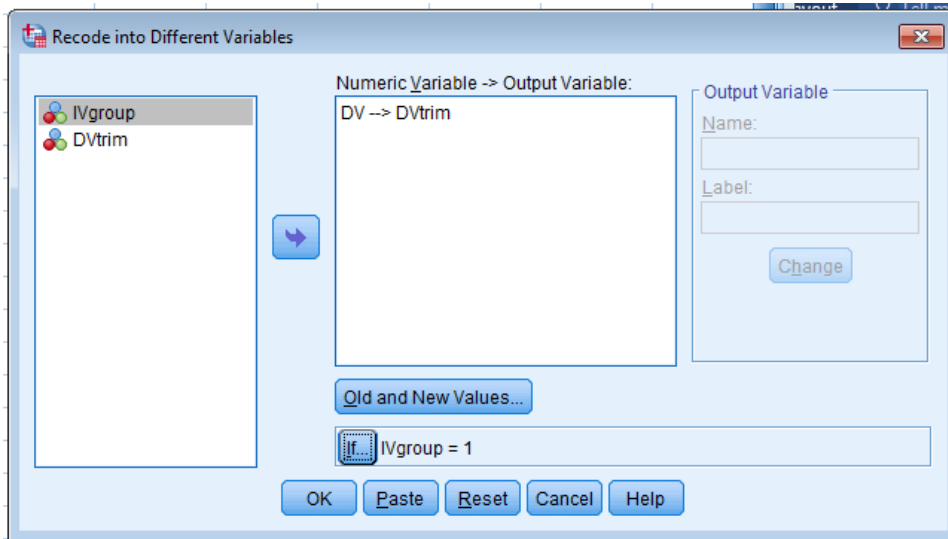
Click on the “Includes if cases...” radio button.

Click on and move the grouping variable into the textbox.

Specify the group you want to work with

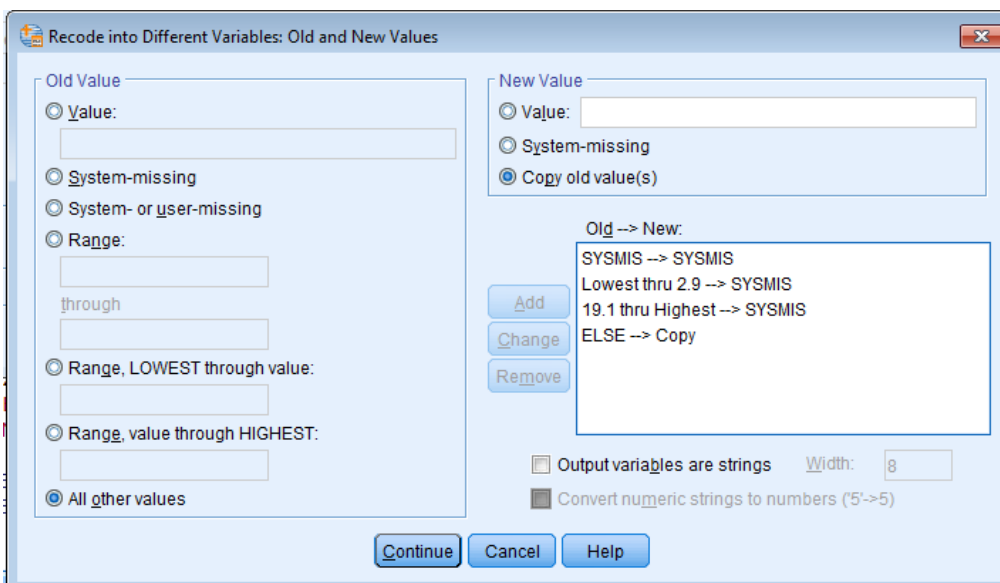
➔ IVgroup = 1

Click on the “Continue” button.



That brings us back to the Recode window.

Now we click “Old and New Values”



Remember to take care of missing values.

Remember that the “boundary values” tell the acceptable values – the values to be Winsorized are anything smaller or larger than those boundaries.

Click “Continue” and then the “OK” on the Main window.

Here is the data set after running this Recode...

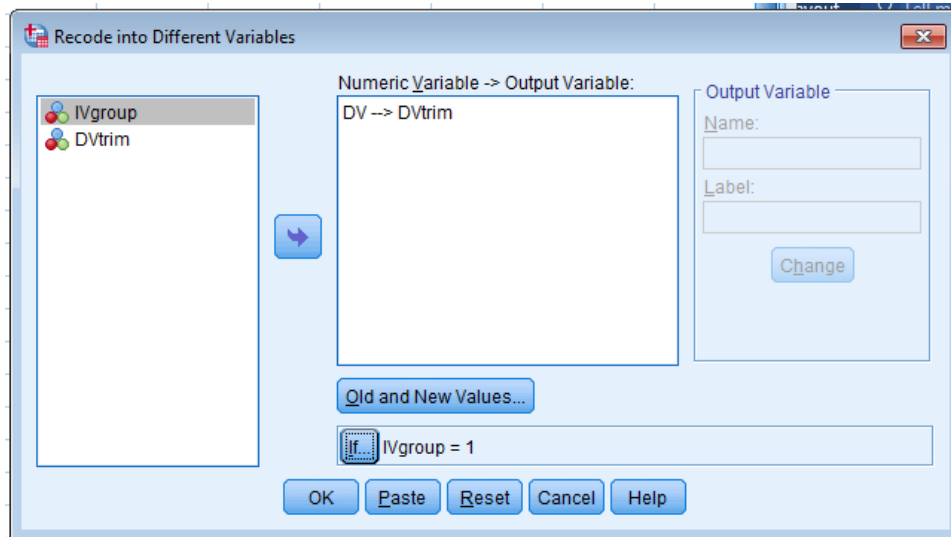
	IVgroup	DV	DVtrim
1	1.00	9.00	9.00
2	1.00	11.00	11.00
3	1.00	7.00	7.00
4	1.00	12.00	12.00
5	1.00	20.00	.
6	1.00	13.00	13.00
7	2.00	20.00	.
8	2.00	29.00	.
9	2.00	28.00	.
10	2.00	26.00	.
11	2.00	34.00	.
12	2.00	27.00	.

Notice that all the IVgroup = 2 cases have missing values. That’s because we only recoded for the IVgroup = 1 cases (using the If option).

So far so good...

The only outlier was the too-large outlier value of 20 which was changed to “missing”.

## Now we do Group = 2

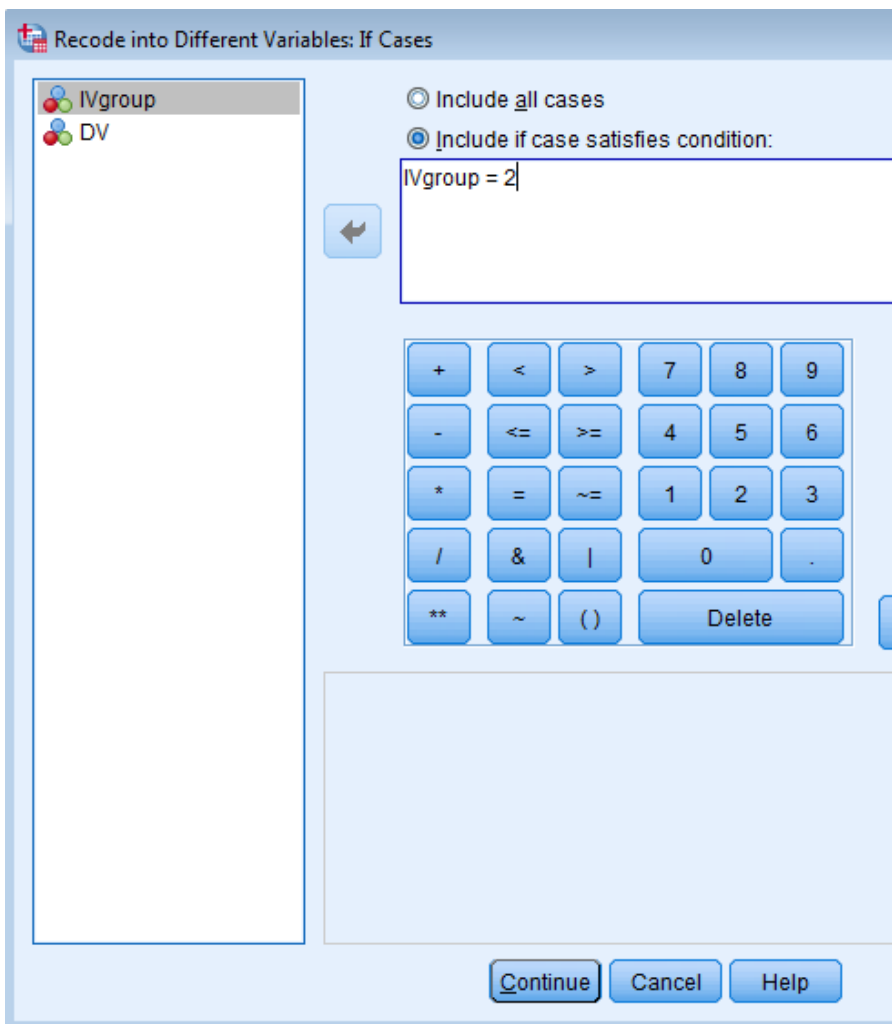


Reopen the Recode in to Different Variables.

It will show you what you ran last time.

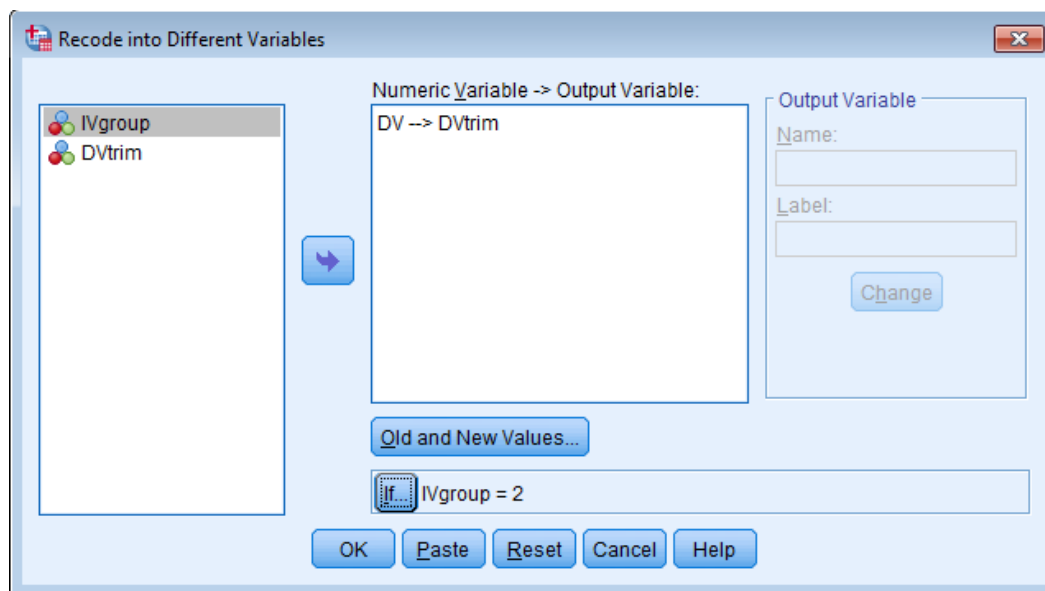
We have to change each part of the specifications  
→ but we use the same variable (DVTrim)!!

Again, click on "If.."



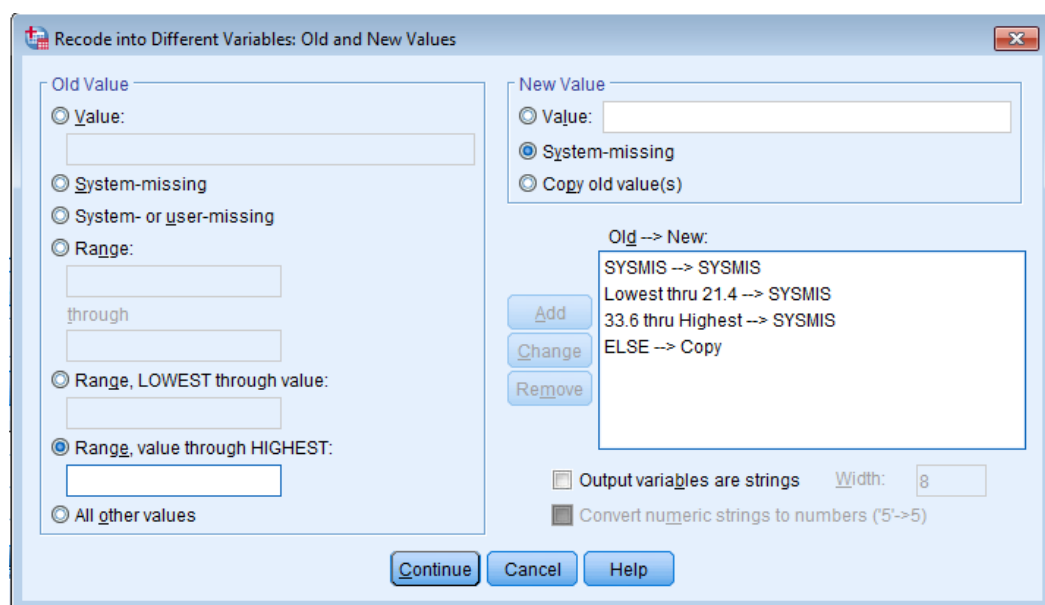
Specify the group you want to work with

→ IVgroup = 2



That brings us back to the Recode window.

Now we click “Old and New Values” so we can change the boundary values to those for Group = 2.



Remember to take care of missing values.

Remember that the “boundary values” tell the acceptable values – the values to be trimmed are anything smaller or larger than those boundaries.

Click “Continue” and then the “OK” on the Main window.

	IVgroup	DV	DVtrim
1	1.00	9.00	9.00
2	1.00	11.00	11.00
3	1.00	7.00	7.00
4	1.00	12.00	12.00
5	1.00	20.00	.
6	1.00	13.00	13.00
7	2.00	20.00	.
8	2.00	29.00	29.00
9	2.00	28.00	28.00
10	2.00	26.00	26.00
11	2.00	34.00	.
12	2.00	27.00	27.00

Now we have trimmed data for both groups.

For Group = 2 we had two outliers 20 was “too small” and 34 was “too large”.

Both of these values have been recoded to “missing”.

Here are the univariates for the raw and the trimmed data:

## Oneway

### Descriptives

DV	N	Mean	Std. Deviation
1.00	6	12.0000	4.47214
2.00	6	27.3333	4.54606
Total	12	19.6667	9.08879

## Oneway

### Descriptives

DVtrim	N	Mean	Std. Deviation
1.00	5	10.4000	2.40832
2.00	4	27.5000	1.29099
Total	9	18.0000	9.20598

One of the disadvantages of trimming is that discarding outliers reduces the sample size and statistical power of the sample.

Group 1 had one “too large” outlier; so trimming it produced a small decrease in the mean for that group.

Group 2 had both a “too small” and a “too large” outlier, so trimming had little effect on the mean.

Both groups have smaller standard deviations after trimming, more so for Group 2 which had 2 values Winsorized.

Here is the syntax code to perform these recodes is:

```
DO IF (IVgroup = 1).
```

```
RECODE DV  
(MISSING=SYSMIS)  
(Lowest thru 2.9=SYSMIS)  
(19.1 thru Highest=SYSMIS)  
(ELSE=Copy)  
INTO DVtrim.
```

```
END IF.  
EXECUTE.
```

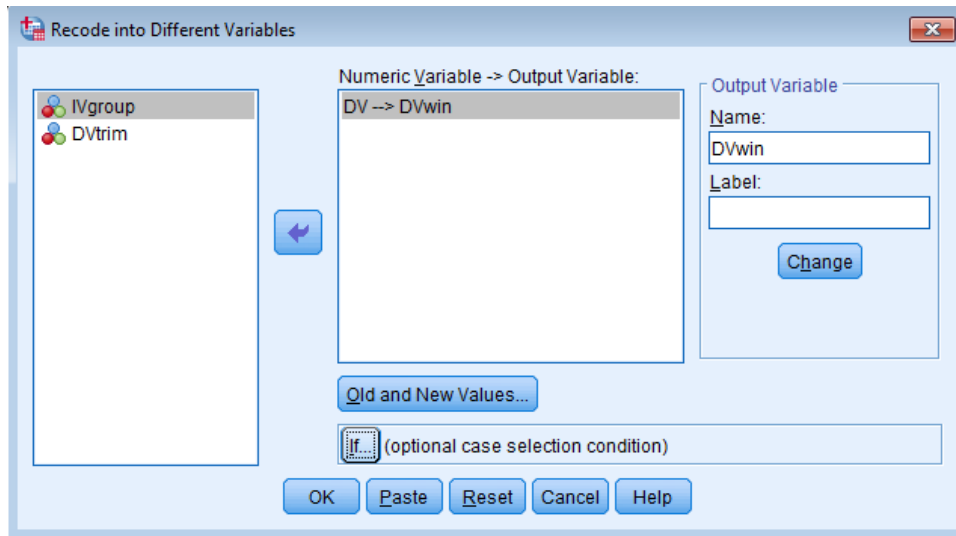
```
DO IF (IVgroup = 2).
```

```
RECODE DV  
(MISSING=SYSMIS)  
(Lowest thru 21.5=SYSMIS)  
(33.6 thru Highest=SYSMIS)  
(ELSE=Copy)  
INTO DVtrim.
```

```
END IF.  
EXECUTE.
```

## Winsorizing

Again, the potentially confusing part is that we have to perform the Winsorizing separately for each group, because the two groups have different outlier boundaries. We will do two separate “runs” using different groups, but using the same variable.



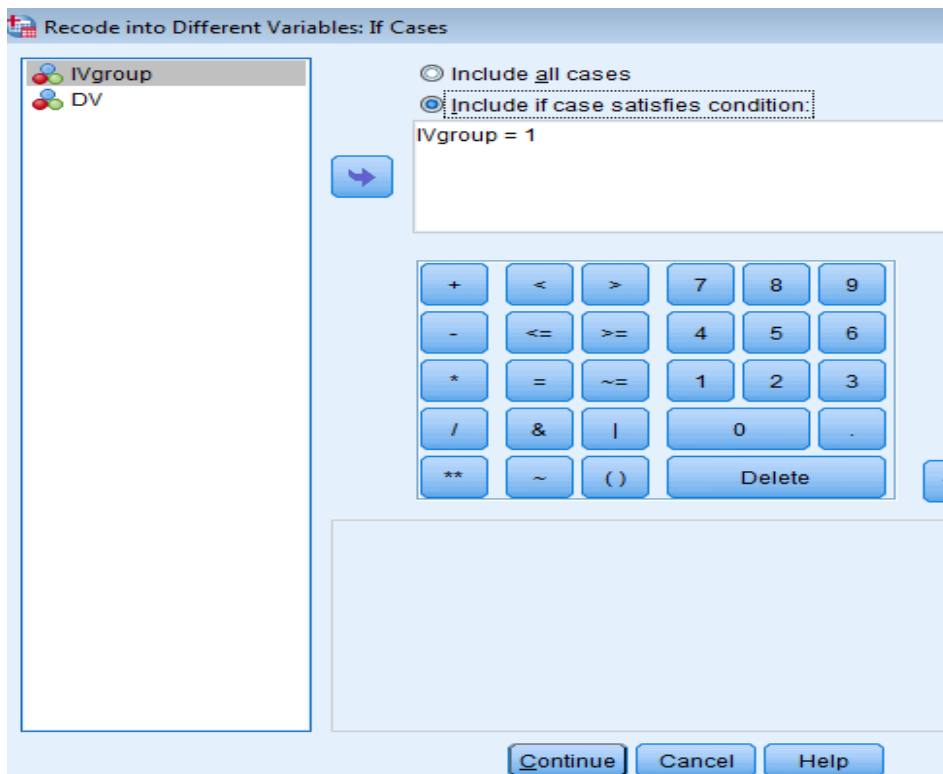
Move the Numeric variable into the textbox.

Specify the name of the Output variable.

Usually we would next click on “Old and New Values”...

But instead, we are going to click on “If...”

## We'll do Group = 1 first...



Click on the “Includes if cases...” radio button.

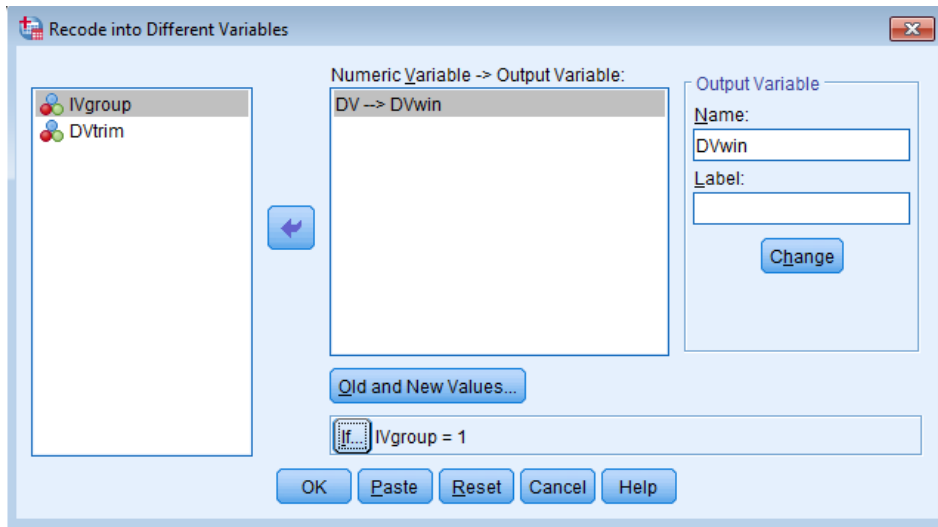
Click on and move the grouping variable into the textbox.

Specify the group you want to work with

➔ IVgroup = 1

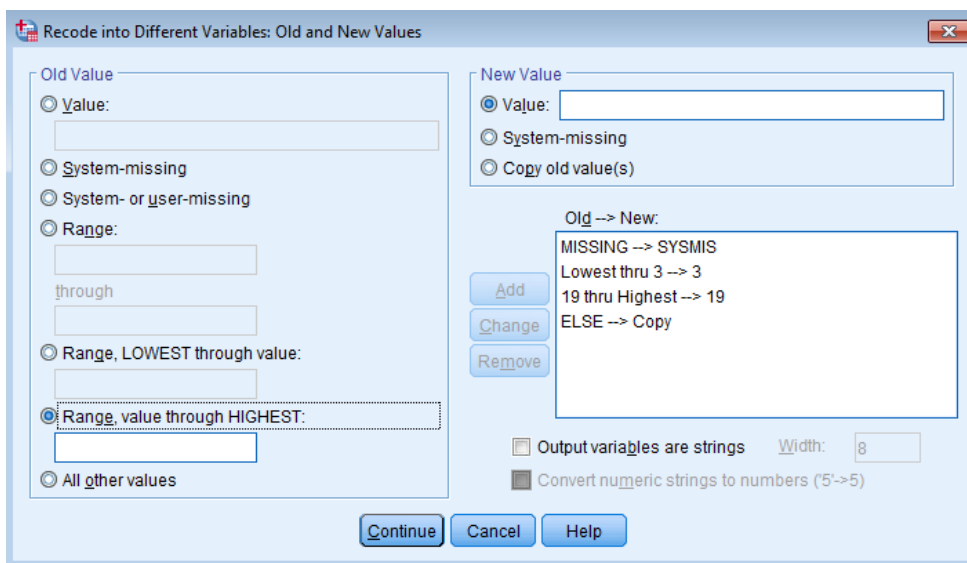
Click on the “Continue” button.





That brings us back to the Recode window.

Now we click “Old and New Values”



Remember to take care of missing values.

Remember that the “boundary values” tell the acceptable values – the values to be trimmed are anything smaller or larger than those boundaries.

Click “Continue” and then the “OK” on the Main window.

Here is the data set after running this Recode...

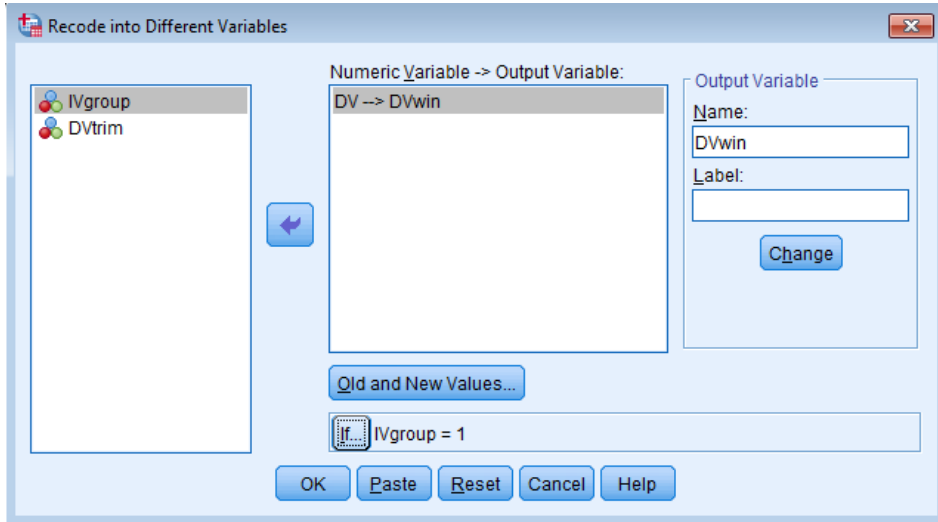
	IVgroup	DV	DVwin
1	1.00	9.00	9.00
2	1.00	11.00	11.00
3	1.00	7.00	7.00
4	1.00	12.00	12.00
5	1.00	20.00	19.00
6	1.00	13.00	13.00
7	2.00	20.00	.
8	2.00	29.00	.
9	2.00	28.00	.
10	2.00	26.00	.
11	2.00	34.00	.
12	2.00	27.00	.

Notice that all the IVgroup = 2 cases have missing values. That’s because we only recoded for the IVgroup = 1 cases (using the If option).

So far so good...

The only outlier was the too-large outlier value of 20, which was changed to “19”.

## Now we do Group = 2

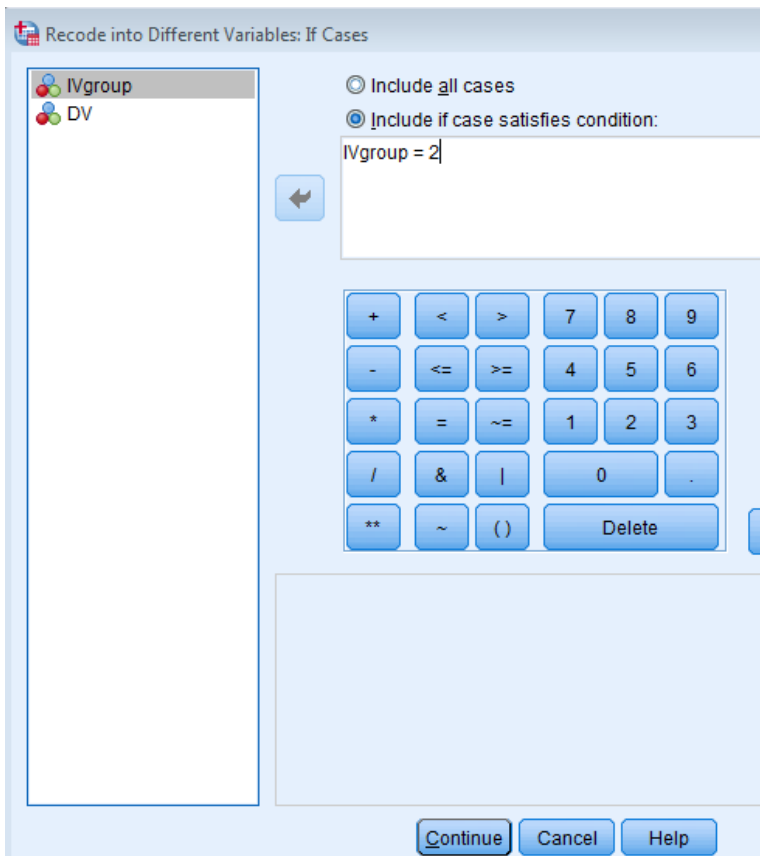


Reopen the Recode in to Different Variables.

It will show you what you ran last time.

We have to change each part of the specifications  
→ but we use the same variable (DVTrim)!!

Again, click on "If.."



Specify the group you want to work with

→ IVgroup = 2

The dialog box 'Recode into Different Variables' shows the variable 'DV' being recoded into 'DVwin'. The 'Old and New Values...' button is highlighted, indicating the next step in the process.

That brings us back to the Recode window.

Now we click “Old and New Values” so we can change the boundary values to those for Group = 2.

The 'Old and New Values' sub-dialog box shows the configuration for recoding. Under 'Old Value', 'Range, value through HIGHEST' is selected. Under 'New Value', 'Value' is selected. The 'Old -> New' list shows the mapping: 'SYSMIS -> SYSMIS', 'Lowest thru 21.5 -> 21.5', and '33.5 thru Highest -> 33.5'. The 'Continue' button is highlighted.

Remember to take care of missing values.

Remember that the “boundary values” tell the acceptable values – the values to be trimmed are anything smaller or larger than those boundaries.

Click “Continue” and then the “OK” on the Main window.

	IVgroup	DV	DVwin
1	1.00	9.00	9.00
2	1.00	11.00	11.00
3	1.00	7.00	7.00
4	1.00	12.00	12.00
5	1.00	20.00	19.00
6	1.00	13.00	13.00
7	2.00	20.00	21.50
8	2.00	29.00	29.00
9	2.00	28.00	28.00
10	2.00	26.00	26.00
11	2.00	34.00	33.50
12	2.00	27.00	27.00

Now we have trimmed data for both groups.

For Group = 2 we had two outliers:

20 was “too small” and was changed to “21.5”

The 34 was “too large” and was changed to “33.5”

Here are the univariates for the raw and the Winsorized data:

## Oneway

### Descriptives

DV	N	Mean	Std. Deviation
1.00	6	12.0000	4.47214
2.00	6	27.3333	4.54606
Total	12	19.6667	9.08879

## Oneway

### Descriptives

DVwin	N	Mean	Std. Deviation
1.00	6	11.8333	4.11906
2.00	6	27.5000	3.92428
Total	12	19.6667	9.03612

One of the advantages of Winsorizing is that it retains all the cases.

Group 1 had one “too large” outlier; so Winsorizing it produced a small decrease in the mean for that group.

Group 2 had both a “too small” and a “too large” outlier, so Winsorizing had little effect on the mean.

Both groups have smaller standard deviations after Winsorizing, more so for Group 2 which had 2 values Winsorized.

Here is the syntax code to perform these recodes is:

```
DO IF (IVgroup = 1).
```

```
RECODE DV  
(MISSING=SYSMIS)  
(Lowest thru 3=3)  
(19 thru Highest=19)  
(ELSE=Copy)  
INTO DVwin.
```

```
END IF.  
EXECUTE.
```

```
DO IF (IVgroup = 2).
```

```
RECODE DV  
(MISSING=SYSMIS)  
(Lowest thru 21.5=21.5)  
(33.5 thru Highest=33.5)  
(ELSE=Copy)  
INTO DVwin.
```

```
END IF.  
EXECUTE.
```