

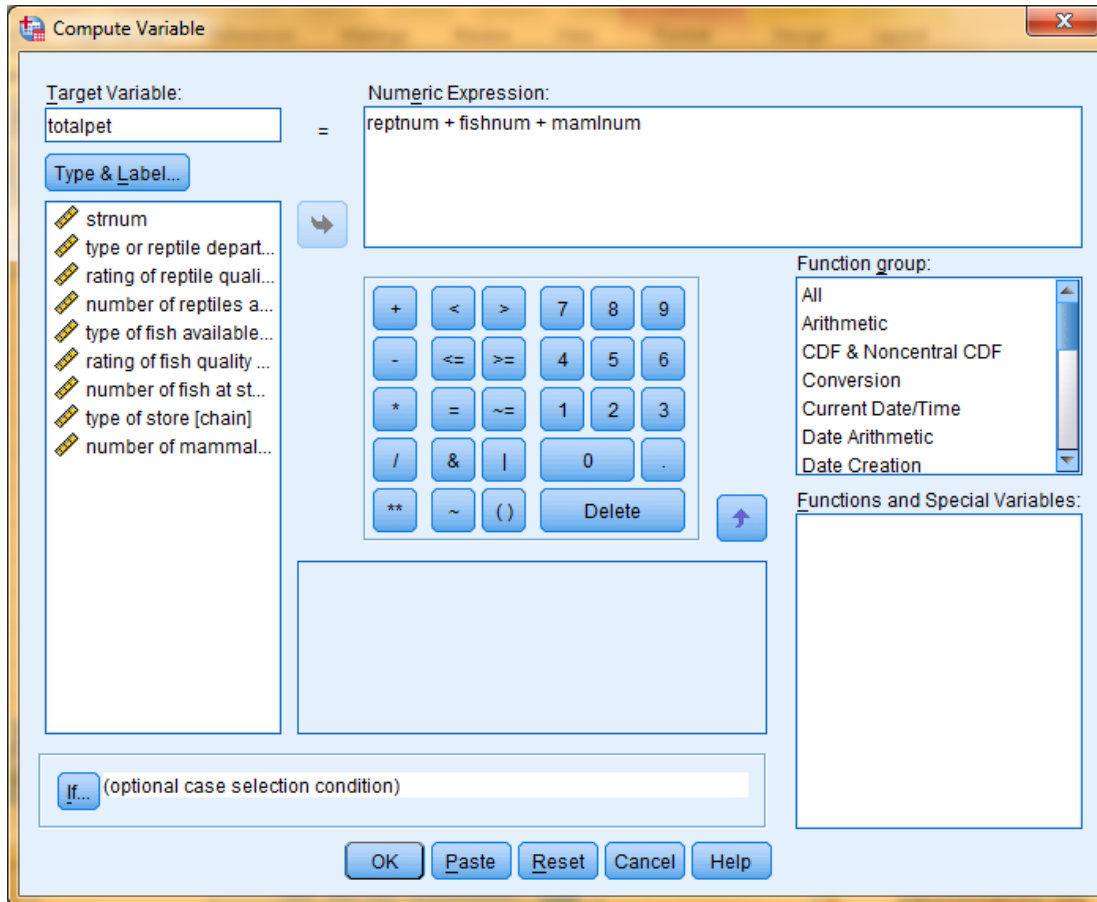
SPSS: Common Data Transformations & Case Management

Using Compute to create a new variable from a formula

From the Data, Syntax or Output windows... **Transform** → **Compute**

- compute statements can be used to create new variables
- the new variable will “appear” in the rightmost column in the data display
- be sure to put in Variable Labels, Value Labels and Missing Values for the new variable
- if a compute statement includes a variable with a missing value for a case, then the resulting variable also has a missing value

Example: We want to compute “totalpet” as the sum of the number of reptiles, fish and mammals



The name of the new variable goes in the “Target Variable” window

Put the formula for the new variable in the “Numeric Expression” window

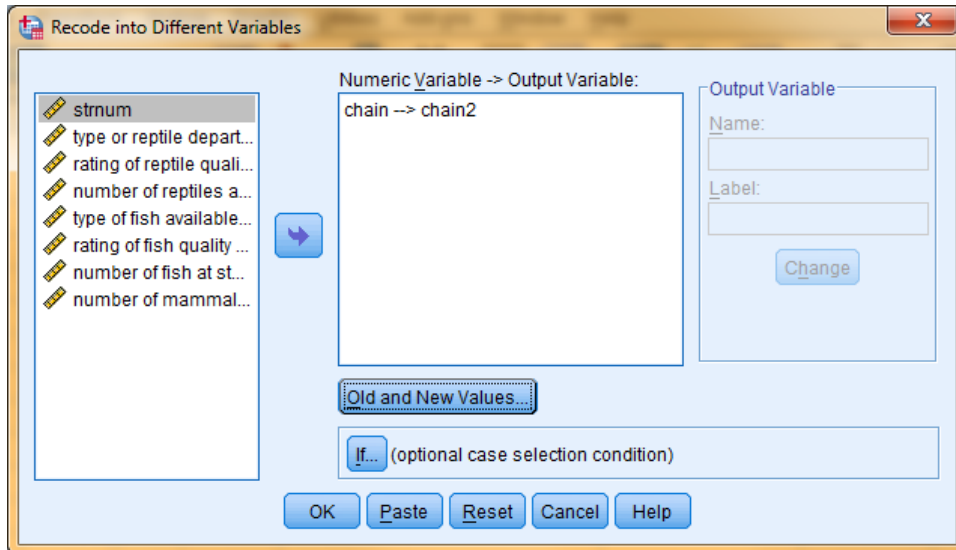
- you can either type in the formula or use the point-n-click buttons
- the “Functions” do a variety of common transformations (dig around a bit)

Using Recode to combine categories of a qualitative/categorical variable

From the Data, Syntax or Output windows... Transform → Recode → Into Different Variables

- recode statements can be used to create new “versions” of variables
- the new variable will “appear” in the rightmost column in the data set
- be sure to put in Variable Labels, Value Labels & Missing Values for the new variable
- if a recode statement includes a variable with a “system missing” value (“.”) for a case, then the resulting variable will automatically have a missing value (.)
- if a recode statement includes a variable with a “user-missing” value for a case, then you must specify that “System- or user-missing values” should become a “System-missing value”

Example: The “chain” variable has three categories. We want to create a new variable “Chain2” that combines chain and coop stores into one category and keeps private stores as a second, separate category

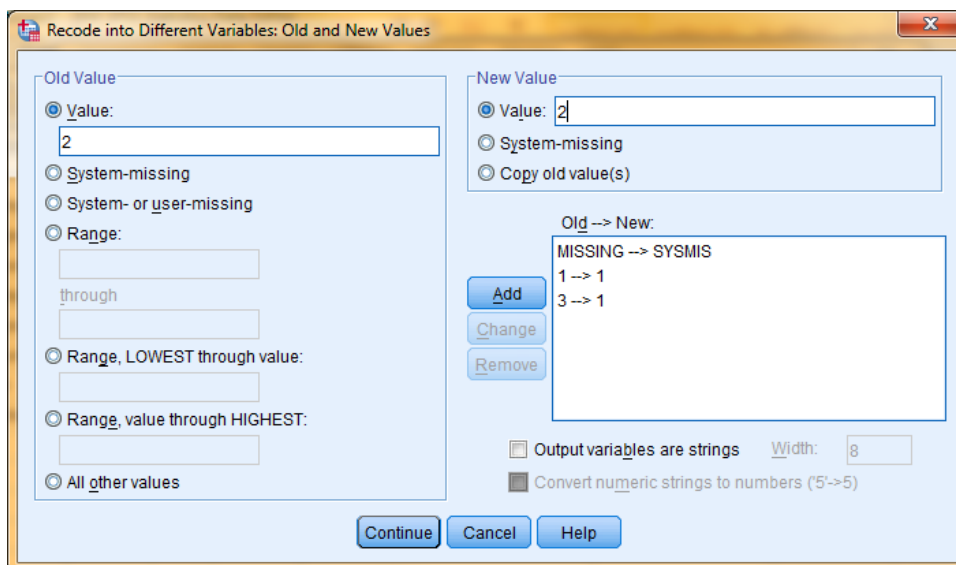


Highlight “chain” in the list of variables and click the arrow to move into the middle window

Type “chain2” into the “Name” box under “Output Variable”

Click the “Change” button

Click the “Old and New Values” button to get the next window



Taking care of missing values

- On the “Old Value” side -- click the radio button beside “System- or User-Missing”
- On the “New Value” side -- click the radio button beside “System Missing”
- click “Add”

Specifying recode values

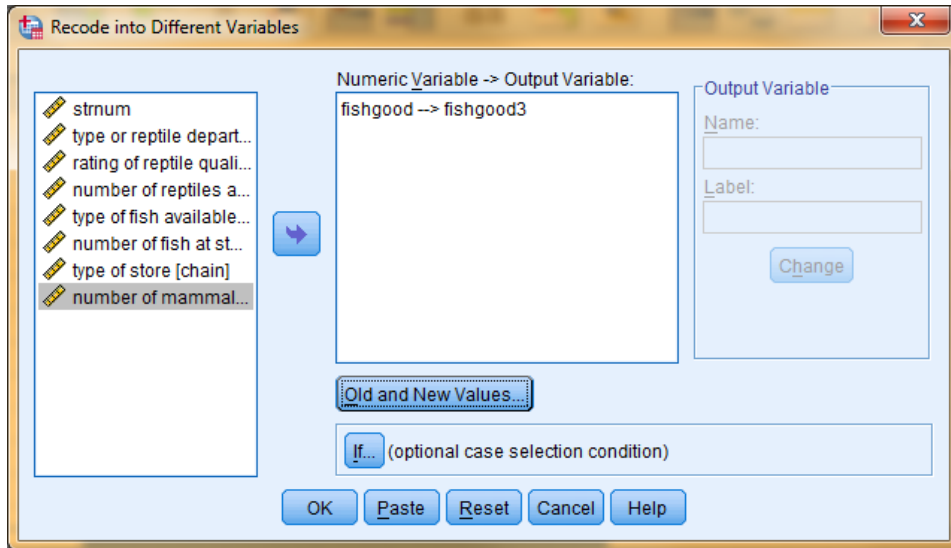
- On the “Old Value” side -- click the radio button beside “Value” and type a “1” into the window
- On the “New Value” side -- click the radio button beside “Value” & type a “1” into the window
- click “Add”
- Repeat with Old & New values until all recodes are visible in the window, then click “Continue”

Using Recode to make a quantitative variable into a set of ordered categories

From the Data, Syntax or Output windows... **Transform → Recode → Into Different Variables**

- recode statements can be used to create new “versions” of variables
- the new variable will “appear” in the rightmost column in the data set
- be sure to put in Variable Labels, Value Labels & Missing Values for the new variable
- if a recode statement includes a variable with a “system missing” value (“.”) for a case, then the resulting variable will automatically have a missing value (.)
- if a recode statement includes a variable with a “user-missing” value for a case, they you must specify that “System- or user-missing values” should become a “System-missing value”

Example: The “fishgood” variable uses a 10-point rating scale. We want to make a new variable “fishrate3” with values of 1-5 = 1, 6-7 = 2 and 8-10= 3

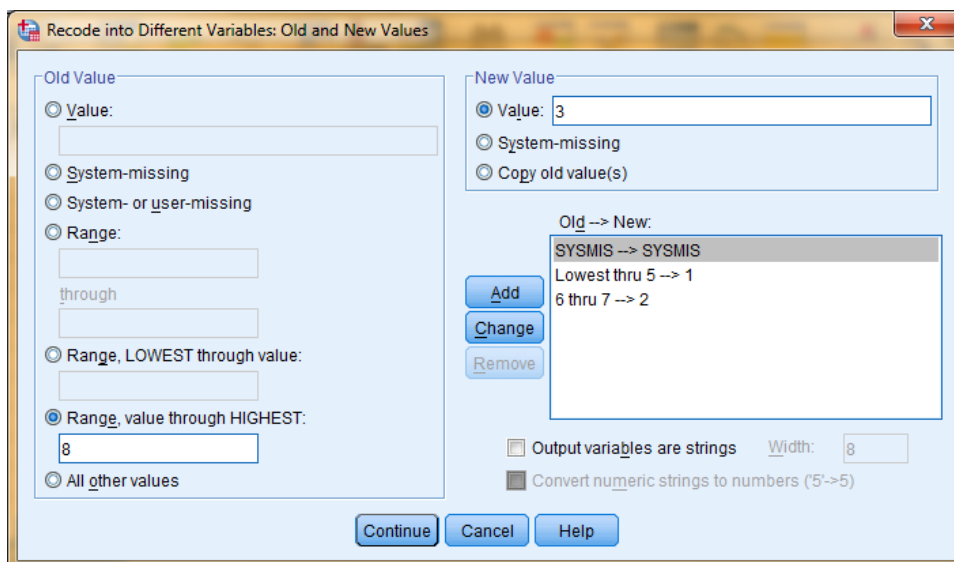


Highlight “fishrate” in the list of variables and click the arrow to move into the middle window

Type “fishrate” into the “Name” box under “Output Variable”

Click the “Change” button

Click the “Old and New Values” button to get the next window



Taking care of missing values

- On the “Old Value” side -- click the radio button beside “System- or User-Missing”
- On the “New Value” side -- click the radio button beside “System Missing”
- click “Add”

Specifying recode values

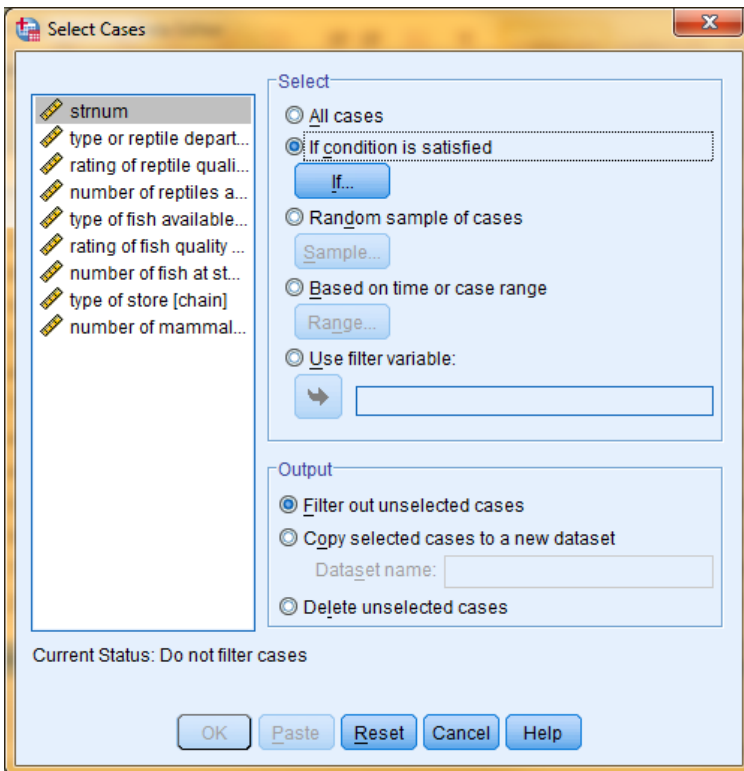
- On the “Old Value” side -- click the radio button beside “Value” and type a “1” into the window
- Use the “Range buttons” to specify what range of the original variable get recoded into the new variable value
- click “Add”
- Repeat with Old & New values until all recodes are visible in the window, then click “Continue”

Data Selection

Sometimes we want to work with only a portion of the data.

Data → Select Cases

Example: We want to analyze just the data from stores that had separate reptile departments (reptdept=2)



Click the “If condition is satisfied” radio button, then click the “If” key.

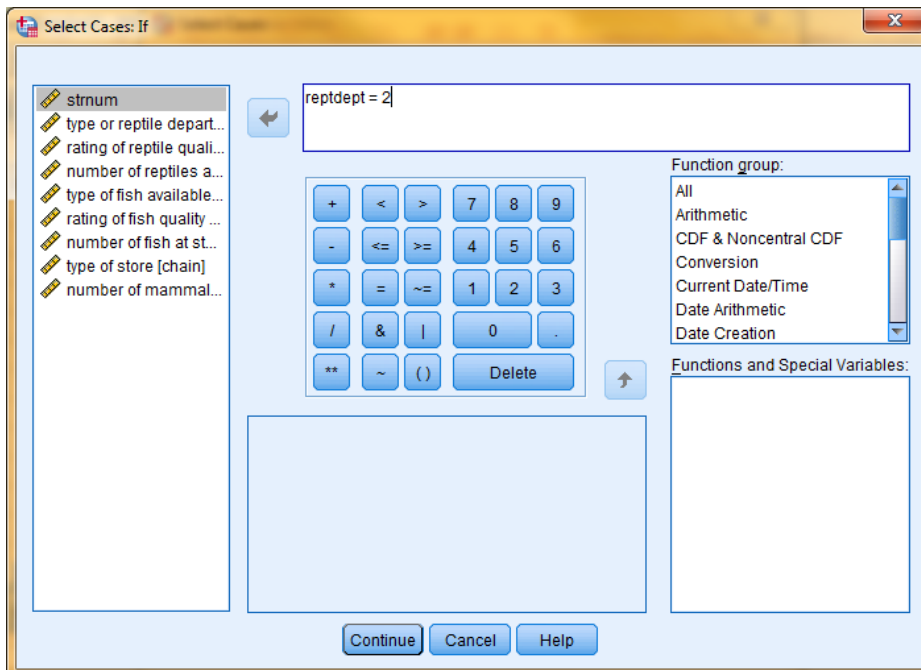
Highlight the variable name and click the arrow to move it into the window.

Use the buttons or type the conditional statement into the window

Click “Continue”

SPSS will create a new variable and place it to the far right in the data set. This new variable will be used to identify and select only those cases that satisfy the conditional for all subsequent analyses.

Be sure to return to the “Select Cases” window and click the “All Cases” radio button when you want to go back to analyzing all the data !!!!!

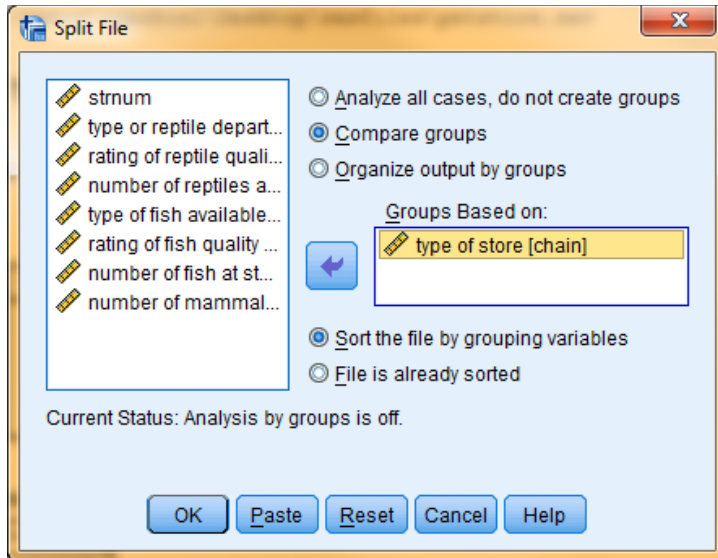


Split Files

Sometimes we want to get separate analyses for each of the “subsets” or “subgroups” in our data.

Example: We want to analyze the data for the three kinds of stores (coop, chain & private) separately.

Data → Split File



Click either the “Compare groups” or “Organize output by groups” radio button. Which one do you want???

Most analyses have multiple parts to the output.

- “Compare groups” will present the first part of the output for all the groups, then the second part of the output for all the groups, etc.
- “Organize output by groups” will present all the output for one group, then all the out for the second group, etc.

Highlight the qualitative variable that defines the subgroups or split files and click on the arrow to move it into the “Groups Based on:” window.

Click “OK”

SPSS will create a new variable and place it to the far right in the data set. This new variable will be used to identify and split the file for subsequent analyses.

Be sure to return to the “Split Files” window and click the “Analyze all cases” radio button when you want to go back to analyzing all the data together !!!!!

Using the SPSS Syntax Window for Data Transformation & Case Management

In addition to the Menus, SPSS has a Syntax Window from which you can perform data transformations, case management and complete statistical analyses using “syntax” or “code”.

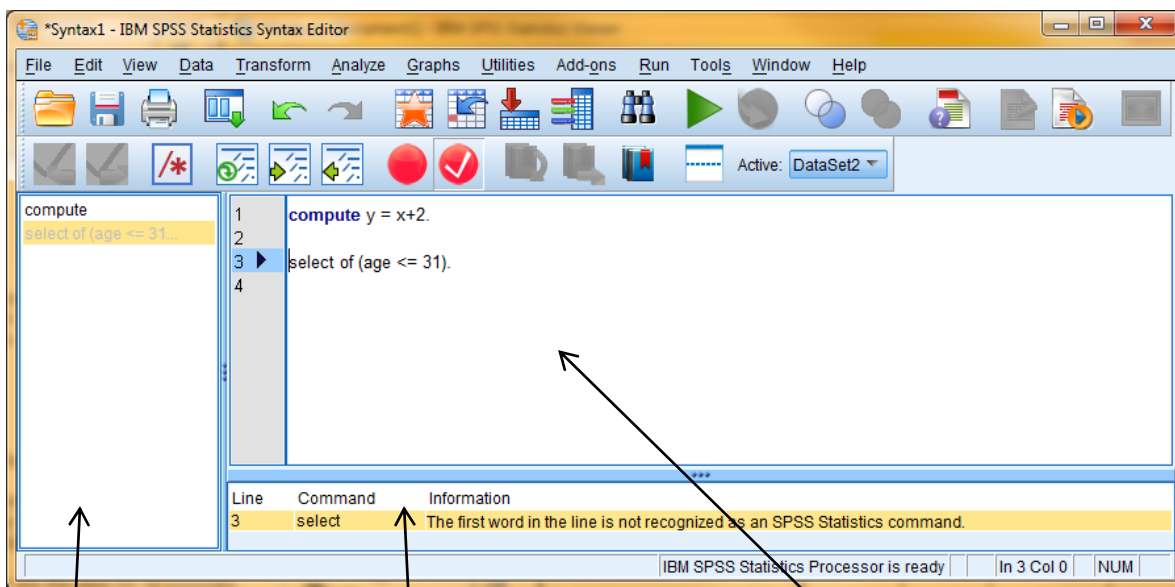
When using the Syntax Window, you have to be careful to use “/” and “.” symbols correctly. Basically, “/” is used to separate the parts of a command and “.” is put at the end of each command.

Missing values can get tricky when doing this. Be sure to proof new variable values – to be sure the transformations worked as you intended!

In general for this and other handouts, the syntax code in **CAPITAL LETTERS** gets typed exactly as it is shown to you, and the parts that are in **lower case letters** represents variable names that will change with each data set and analysis. However, in practice, either capital or lower case letters will work for most commands.

You also need to understand the use of the “exe.” or “execute.” command. When you type in the syntax for a data transformation (e.g., using a “Compute” or “Recode” statement) and run that command, SPSS creates the new variable and puts a column for it on the right hand side of the database. But, if you look at that column, it will be filled with “.” Indicating missing values for the variable. That is because the new data values have not actually been computed. The values will be computed when: 1) you request a statistical analysis (whether or not it uses the new variables) or 2) you include and run an “exe.” command after the transformation or case management code,

To get a syntax window use **File → New → Syntax**



The left frame keeps track of the different commands you have entered into the syntax work space, so that you can navigate among them.

The syntax log gives feedback about programming error,

Errors are also reported in the output window.

The main work space of the Syntax window. As you type commands it generates line numbers, which makes it easier to understand error messages, etc,

Syntax windows can be saved (File → Save). The complete information in all three frames is saved (with the extension “.sps” and can be opened and used or modified.

SPSS Syntax to COMPUTE a new variable

- Type the compute statement(s) with the new variable and the formula into the SPSS Syntax Window
- Highlight the statement(s)
- Click on the big green (run) arrow near the top of the window (under "Tools")

COMPUTE totalpet = reptnum + fishnum + mamlnum. ← computes total pet as the sum of the three variables

COMPUTE perf_prop = numcorrect / 25. ← computes the proportion performance score of 25 items

COMPUTE perf_per = (numcorrect / 25) * 100. ← computes the percentage performance score of 25 items

*it is a good idea to provide Variable Labels for created variables. ← comment

VARIABLE LABEL perf_prop "Proportion correct on performance test" ← variable name with label in parentheses
 perf_per "Percentage correct on performance test".

EXE.

SPSS Syntax to combine categories of a qualitative variable using RECODE

Example: The "chain" variable has three categories. We want to create a new variable "Chain2" that combines chain and coop stores into one category and keeps private stores as a second, separate category

RECODE chain ← name of the starting variable
 (MISSING=SYSMIS) (1=1) (3=1) (2=2) ← specify values -- in the order (original=recoded)
 INTO chain2. ← name of resulting variable

*It is a good idea to provide a Variable Label and a Value Label for created variables. ← comment

VARIABLE LABEL chain2 "Type of Store recoded". ← name of resulting variable with label in parentheses

VALUE LABELS chain2 ← name of the resulting variable
 1 "Chain & Coop Stores" ← each variable value with the label in parentheses
 2 "Private Stores".

EXE.

SPSS Syntax to make a quantitative variable into a set of ordered categories using RECODE

Example: The "fishgood" variable uses a 10-point rating scale. We want to make a new variable "fishrate3" with values of 1-5 = 1, 6-7 = 2 and 8-10 = 3

RECODE fishgood ← name of starting variable
 (MISSING=SYSMIS) (Lowest thru 5=1) (6 thru 7=2) (8 thru Highest=3) ← specify values (original=recoded)
 INTO fishgood3. ← name of resulting variable

*It is a good idea to provide a Variable Label and a Value Label for created variables. ← comment

VARIABLE LABEL fishgood3 "recoding of fishgood". ← name of resulting variable with label in parentheses

VALUE LABELS fishgood3 ← name of the resulting variable
 1 "Poor quality" 2 "Fair Quality" 3 "Good Quality". ← each variable value with the label in parentheses

EXE.

SPSS Syntax for Conditional COMPUTE & RECODE Statements

Sometimes you may want to apply different scoring or coding schemes for different subsets of cases.

Conditional COMPUTE Statements

Example: For those stores with separate reptile departments (reptdept = 2) we want to compute total pet as the sum of reptiles, fish and mammals, whereas for those stores that do not have a separate reptile department (reptdept = 1), we want to compute the total pets as the sum of fish and mammals.

```
IF (reptdept = 1) totalpet2 = fishnum + reptnum + mamlnum. ← will only compute totalpet2 for those with "reptdept = 2"  
IF (reptdept = 2) totalpet2 = fishnum + mamlnum. ← will only compute totalpet2 for those with "reptdept = 1"
```

```
VARIABLE LABEL totalpet2 "total pet computed including reptnum only for stores with separate reptile departments".
```

Conditional RECODE Statements

Example: For stores with separate reptile departments (reptdept = 2) we want to make ratings of 1-5=1, 6-7=2 and 8-10=3, whereas for those stores that do not have separate reptile departments (reptdept= 1) we want to make ratings of 1-6=1, 7-8=2 and 9-10=3

```
DO IF (reptdept = 1).  
RECODE fishgood (MISSING=SYSMIS) (Lowest THRU 6 = 1) (7 THRU 8 = 2) (9 THRU Highest = 3) INTO fishgood4.  
END IF.
```

```
DO IF (reptdept = 2).  
RECODE fishgood (MISSING=SYSMIS) (Lowest THRU 5 = 1) (6 THRU 7 = 2) (8 THRU Highest = 3) INTO fishgood4.  
END IF.
```

```
VARIABLE LABEL fishgood "fish quality rating computed differently for stores with separate reptile department".
```

```
VALUE LABELS fishgood  
1 "Poor quality" 2 "Fair Quality" 3 "Good Quality".
```

```
EXECUTE.
```