Some basics…
Creating variables is useful for discovering information about your dataset.

Important examples of data manipulation include, creating dummy variables, interaction terms, and expanding functional form assumptions.

Now: Import the file affairs from the desktop into SAS (it is in tab delimited format)

Binary variables
Dummies (Binary variables)

- Dummy variable creation is relatively easy in SAS

Example:
Binary Variables

HOW?

Data work.affair;
Set affair;
If marrat=2 then unhap=1;
ELSE unhap= 0;
Run;

Repeat for average, happy, and very happy or combined the code into one set of statements

Binary Variables

Now let’s say we want to classify individuals according to educational attainment within the dataset according to the following formulation.

High if education is less then or equal to 12 years
College if education is between 12 and 16
Grad if education is greater then 16.

Viewtable

It is a good Idea to check to make sure your new variables were encoded correctly.

Click on the explorer table
Click on work folder
Click on the viewtable affair

A Spreadsheet of affairs opens up.

BE SURE TO CLOSE VIEWTABLE BEFORE CONTINUEING!!!

If you don’t new variables cannot be encoded!

Binary Variables

The following code could generate the new variables.

Data work.affair;
Set affair;
If educ=<12 then high=1; ELSE high= 0;
If educ>12 and educ<=16 then Col=1;
ELSE Col= 0;
If educ>16 then Grad=1; ELSE Grad= 0;
Run;
Variable encoding

Transforming variables via some operator can be very useful. Operators include:

- * Multiplication
- ** Exponents
- + addition
- / division
- Log() natural logarithm

Transformation

Let's say I want to add education squared to the regression. I can create this variable via the following statement:

\[ \text{Educsq} = \text{educ}^2 \]

I can create the natural logarithm of years married by using:

\[ \text{Lnmar} = \log(\text{yrsmarr}) \]

Interaction Terms

I can create any number of interaction terms by simply putting in both variable names in a formula:

Interaction of years married and good marriage:

\[ \text{Hapyear} = \text{Vryhap} \times \text{yrsmarr} \]

Now: Import the dataset CandY into SAS (it is in tab delimited format)

Time Series Variables

Let's say I have disposable income and want to look at lags of disposable income. How?

```
Data work.candy;
Set candy;
disy_1 = lag(disy);
Run;
```
Time Series Variables

Additional Lags can be created by using the lag operator and \( n \).

For 2 lags you could enter:

```sas
Data work.candy;
Set candy;
disy_2=lag2(disy);
Run;
```

Differences can also be calculated either using formulas:

\[ \text{FDIFY} = Y - Y_{-1} \]

Or by the Dif operator:

```sas
Data work.candy;
Set candy;
FDIFY=dif();
Run;
```

To compute Second Difference the following would be used:

```sas
Data work.candy;
Set candy;
fdf=Dif(Dif(disy));
Run;
```

The difference between a variable and its \( n \)th lag can be calculated by supplying:

```sas
newvariable=Difn(argument)
```

If I wanted the difference between disy and its 3 lag you would type:

```sas
Data work.candy;
Set candy;
tdf=DIF3(disy);
Run;
```
Time Series Variables

1. Special Note: Dif2 and Dif(Dif())) are not the same!!
2. In order to use these special functions observations must be in chronological order!!
3. All new variables created should be checked for accuracy. By examination (or evaluation of means).

The End