Command	**Command Description**
Data Set Creation	
Data "dataset name";	You are naming a data set that you will input and manipulate data. You will refer back to the data name to, for example, merge with other files etc.
Infile "external file location and name"	Typically the raw data you will work with will be in a separate file. This command points SAS to that file.
Input (variable names);	This command directs SAS to input the specific variables within the raw data set designated in the Infile statement. If your raw data is unformatted with spaces between variable, you could use an unformatted input statement. If the data is formatted, the different variables in specific positions, then it's usually better to use formatted statements. For example if the data set specified in you Infile statement looks like: David Sui 500 24 14 James Logan 439 13 18 Mike Len 523 18 12 The formatted input statement for the data: Input @1 Name $ 12. @13 Salary 5. @21 exp 2. @29 ed 2.; The @ sign tells SAS which position the variable starts in. The $ symbol informs SAS that the variable is non-numeric. The first variable takes up a maximum of 12 spaces so you should include 12. in the statement after the variable. The education and experience variables both take up two spaces at most , so they are given 2. after each.

Arithmetic Operators:
    addition (+), subtraction (-),
    muliplication (*), division (/)
	Example: suppose the above salary data were weekly salary for a sample of fulltime workers. I want to create a new variable representing monthly salary: msalary=salary*4; or to create hourly salary: hsalary=salary/40.

Comparison Operator	
Less than (<), less than or equal to (<=)	Example: suppose I wanted to limit my data set only to those with experience levels greater than 15 years, I would include the statement: if exp>15
Greater than (>), greater than or equal to (>=)	Example: Suppose I wanted to create a separate variable (called college) that equals one if the person in the sample has a college degree and zero otherwise: if ed>=16 then college=1; else college=0;
Functions Absolute value of variable X: ABS(x) Natural log of variable X: Log(x)	N'th Lag of Variable: Lagn(X)

Command	**Command Description**

Proc Statements

Proc Print;Var (variable name);	If you include only Proc Print into a data statement without the variable statement, SAS will automatically print all the data within the statement. If you want to designate individual variables to print, for example only the name and experience variables. Include the statement: **var name exp;**
Proc Sort; By (variable name);	This procedure sorts the data by whatever variable you designate. For example to sort our data above by the name of the wage-earner: **Proc Sort;by name;** If you want to merge two data sets by some variable then the data sets must be sorted first. For example, if you have another data set with the sampled wage-earners name and age, you can merge the two data sets into one with all the variables included. The statement: **Merge (first data set) (second data set);by name;**
Proc Means;	Calculates summary statistics for designated variables. Include a Var statement to specify variables. To specify the statistics you want, some of the options are:

n	number of observations
mean	Mean
min	minimum value of data
max	maximum value of data
sum	sum of data values
std	sample standard deviation
t	t-value for the mean =0
prt	prob-value for the t-test

Proc Corr;Var(variable names)	Calculates a matrix of correlations across the paired variables stated in the variable statement.
Proc ttest; Class (variable name); Var (variable name)	Calculates t-test for the difference in means in a variable. The separate means are classified by the class variable. For example, if wage data were identified by male and female workers the gender variable would be the class variable.
Proc Freq;Tables A;	Calculates the frequency of each value of a variable defined as **A.**
Proc Freq;Tables A*B;	Calculates a crosstabulation of the variables designated A and B. The crosstabulation will show the frequency of pairs of the two variables.

Command	**Command Description**

Proc Reg; Model dependent = independent **/** options**;**

Performs regression of the dependent on independent variables. The options within the statement include:

dw	Calculates Durbin-Watson statistic
noint	Specifies no intercept in model

To create a SAS data set with regression model residuals or predicted values, use the commands: **Output Out**=dataset name **Predicted**=variable name **Residual**=variable name;