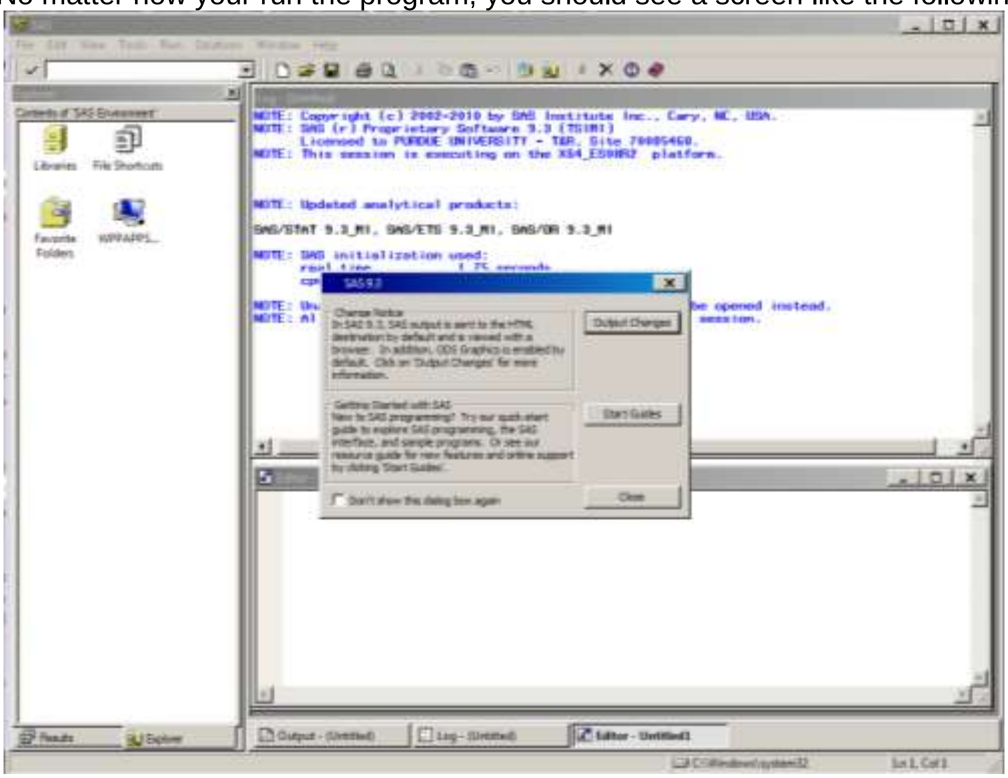
Introduction to SAS

Getting Started

1) Download all of the files, sas programs (.sas) and data files (.dat) into one of your directories. I would suggest using your H: drive if you are using a computer at Purdue or goremote.

2) Run SAS
There are two ways to run SAS on an ITAP computer.
   1) On ITAP Computers: Select All Programs → Standard Software → Statistical Packages → SAS 9.3b.
   2) Using goremote: Go to https://goremote.itap.purdue.edu/Citrix/XenApp/auth/login.aspx and log in using your career account. Find the program by selecting Standard Software → Statistical Packages → SAS 9-3 64-BIT.

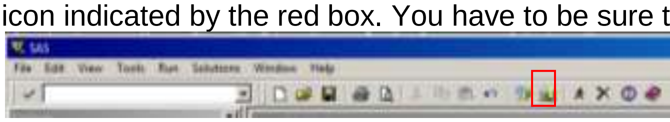
3) No matter how your run the program, you should see a screen like the following:



You may want to look at the Start Guides, but unless you have used a previous version of SAS, looking at the Output Changes will not be beneficial. Click on Close.

4) To load up your sas program, go to File → Open Program and locate and open the program of interest, say start.sas.

5) You will see the program in the Editor window. To run the program click the ‘running man’ icon indicated by the red box. You have to be sure that the Editor window is active.



Getting the data into SAS

There are two ways of getting your data into SAS; you can either enter the data by hand or you can load from a pre-existing file.

Sample code for entering data by hand:

```
*from data in the code this example is Example 2.2.4, liter size of sows;
data litter;
input piglets sows @@;
datalines;
5 1 6 0 7 2 8 3 9 3 10 9 11 8 12 5 13 3 14 2
;
run;

proc print data=litter; run;
```

The colors of the file indicate different things:
blue: keywords from the program
black: items that you enter
green: comments
yellow highlight: inputted data
bluegreen: numbers in code
purple: titles and text

@@: data is in rows instead of columns

Additional notes common to all SAS source codes:

1) Each line always ends in a semicolon, ;
2) 6) You should always have the statement 'quit;' at the end of each script.

Sample code loaded from pre-existing file:

```
*from a file;
data liter2;
infile 'H:\My Documents\Stat 503\pigs.dat';
input piglets sows;
run;
proc print data=liter2;
run;
```

Creating a Histogram

The following code creates the same histogram as in Fig. 2.2.4.

```
data liter3;
infile 'H:\My Documents\Stat 503\Ex.2.2.4.dat';
input piglets;
run;
proc print data=liter3;
run;

title1 'Surviving Piglets';
*creates a histogram using the classes 5, 6, ...., 15;
proc univariate data=liter3 noprint;
histogram piglets / endpoints = 5 to 15 by 1;
run;
```

If you would want to change the classes, you would change what the classes are, you would change the numbers after the keyword ‘endpoints’. The first two numbers indicate the range and the last indicates the the size of the bin. For example, if you would want bins from 10 – 50 with a width of 5, you would type ‘`histogram piglets / endpoints = 10 to 50 by 5;`’

Manipulating data sets

To following code exemplifies how you copy one set to another one and add another variable.

```
data a2;
  set liter3;
  original = 'yes';
run;

proc print data=a2; run;
```

output:

Surviving Piglets

Obs	piglets	original
1	5	yes
2	7	yes
3	7	yes
4	8	yes
5	8	Yes

To combine two data sets, use the following code:

```
data a3;
  piglets = 15; output;
  piglets = 16; output;
run;

data a4;
  set a2 a3;
run;

proc print data=a4; run;
```

This will add the following rows to the above data set:

Obs	piglets	original
37	15	
38	16	

output from univariate

Finally, we have discussed in class the descriptive stattistics, the are in proc univariate. In addition, we can generate the histogram (automatically) and the QQ-Plot (Chapter 4). These last two plots are used to determine if the distribution is normal.

```
proc univariate data=liter3;
   var piglets; /*gives many numerical summaries for the variable piglets*/
   histogram piglets/normal; /*histogram*/
   qqplot piglets; /*qqplot*/
run;
```

Output:

Surviving Piglets

The UNIVARIATE Procedure
Variable: piglets

Moments			
N	36	**Sum Weights**	36
Mean	10.4166667	**Sum Observations**	375
Std Deviation	1.99105141	**Variance**	3.96428571
Skewness	-0.4792508	**Kurtosis**	0.48822726
Uncorrected SS	4045	**Corrected SS**	138.75
Coeff Variation	19.1140935	**Std Error Mean**	0.3318419

Basic Statistical Measures			
Location		**Variability**	
Mean	10.41667	**Std Deviation**	1.99105
Median	10.50000	**Variance**	3.96429
Mode	10.00000	**Range**	9.00000
		Interquartile Range	2.50000

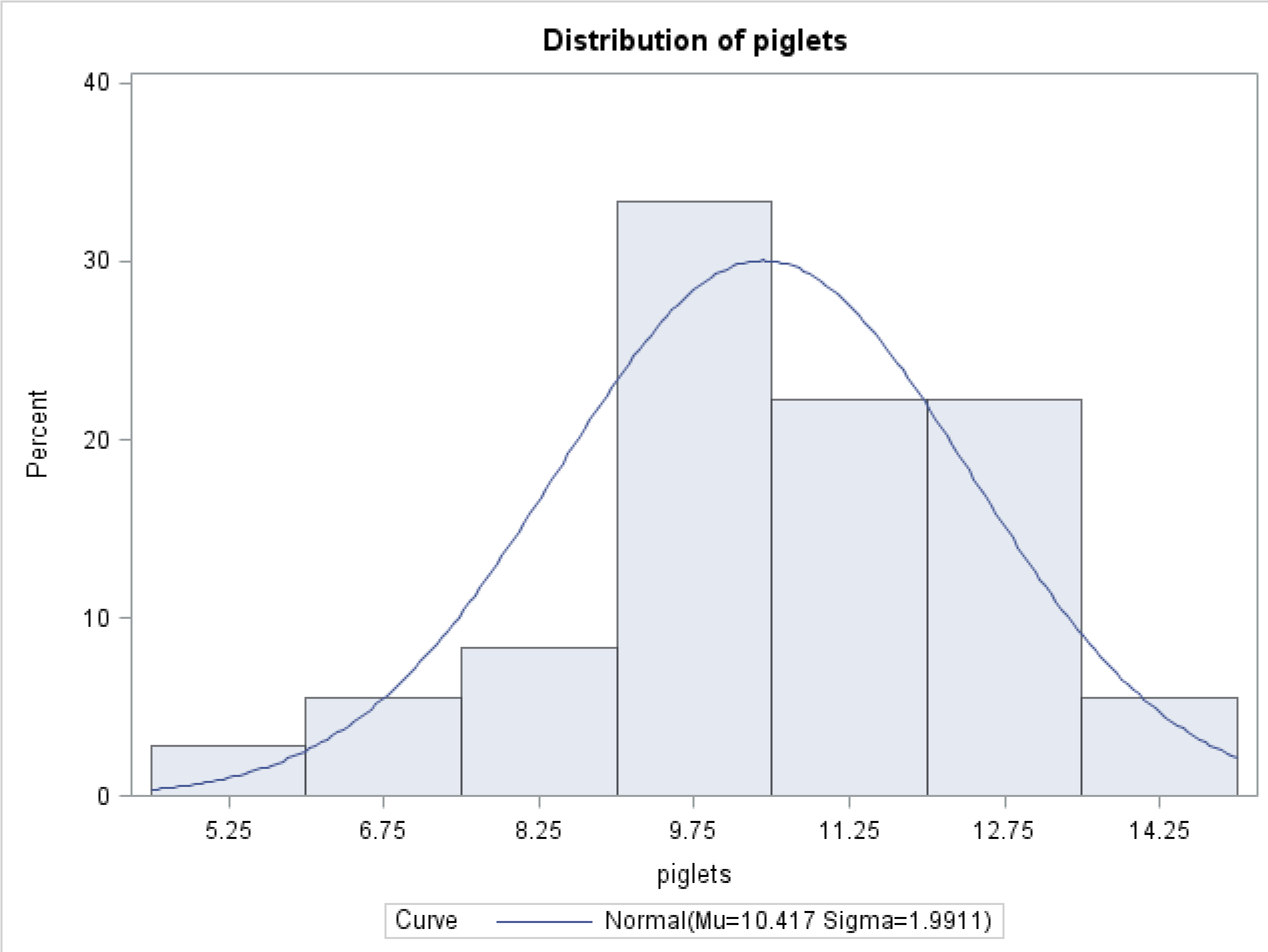
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	31.39045	**Pr > |t|**	<.0001
Sign	**M**	18	**Pr >= |M|**	<.0001
Signed Rank	**S**	333	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Quantile	**Estimate**
100% Max	14.0
99%	14.0
95%	14.0
90%	13.0
75% Q3	12.0
50% Median	10.5
25% Q1	9.5
10%	8.0
5%	7.0
1%	5.0
0% Min	5.0

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
5	1	13	32
7	3	13	33
7	2	13	34
8	6	14	35
8	5	14	36

Surviving Piglets

The UNIVARIATE Procedure



Surviving Piglets

The UNIVARIATE Procedure
Fitted Normal Distribution for piglets

Parameters for Normal Distribution		
Parameter	**Symbol**	**Estimate**
Mean	Mu	10.41667
Std Dev	Sigma	1.991051

Goodness-of-Fit Tests for Normal Distribution				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.16711886	**Pr > D**	0.012
Cramer-von Mises	**W-Sq**	0.11409952	**Pr > W-Sq**	0.073
Anderson-Darling	**A-Sq**	0.59245060	**Pr > A-Sq**	0.118

Quantiles for Normal Distribution		
Percent	**Quantile**	
	Observed	**Estimated**
1.0	5.00000	5.78479
5.0	7.00000	7.14168
10.0	8.00000	7.86503
25.0	9.50000	9.07372
50.0	10.50000	10.41667
75.0	12.00000	11.75961
90.0	13.00000	12.96830
95.0	14.00000	13.69165
99.0	14.00000	15.04854

Surviving Piglets

The UNIVARIATE Procedure