```
* Introduction to SAS;
* Copyright 2013 by Ani Katchova;

* Creating a library where SAS data files are located;
libname lib1 'C:\Econometrics\Data';
run;

* Using a sas file;
data auto;
set lib1.intro_auto;
run;

* Importing a csv file;
proc import out=auto_csv
datafile = "C:\Econometrics\Data\intro_auto.csv"
dbms=csv replace; getnames=yes; datarow=2;
run;

* Importing an Excel file;
proc import out=auto_excel
datafile = "C:\Econometrics\Data\intro_auto.xlsx"
dbms=excel replace; getnames=yes;
run;

* Print data in output window;
proc print data=auto (obs=10);
run;

* Listing the variables;
proc contents data=auto;
run;

* Sorting the data;
proc sort data=auto;
by make price;
run;

* Descriptive statistics;
proc means data=auto;
*class foreign;
run;

* Detailed descriptive statistics;
proc univariate data=auto;
var price mpg;
run;

* Descriptive statistics saved in a file;
proc sort data=auto; by make; run;
proc univariate data=auto noprint;
var price mpg;
by make;
output out=stats mean=price_mean mpg_mean std=price_std n=obs;
run;
```

```
* Merging two files - "make" is the common variable;
proc sort data=auto; by make; run;
proc sort data=stats; by make; run;
data auto2;
merge auto stats;
by make;
run;

* Frequency distribution for different variables;
proc freq data=auto;
tables make foreign repairs;
run;

* Bar chart;
proc gchart data=auto;
vbar make/ discrete;
run;

* Correlations;
proc corr data=auto;
var price mpg weight length;
run;

* Correlations by group - need to sort first and then use "by" statement;
proc sort data=auto; by foreign; run;
proc corr data=auto;
var price mpg weight length;
by foreign;
run;

* Data manipulations;
data auto1;
set auto;
if price>6000 then highprice=1; else highprice=0;
if repairs=0 or repairs=1 then repairs_level=0; else if repairs=2 or repairs=3 then
repairs_level=1; else repairs_level=2;
price_nominal=price*3.35;
if price=. then delete;
price_per_weight=price/weight;
where foreign=0;
run;

* Regression model - mpg is dependent variable and weight, length and foreign are
independent variables;
proc reg data=auto;
model mpg = weight length foreign;
run;

* ANOVA test - if the mean mpg is the same for foreign and domestic cars;
proc glm data=auto;
class foreign;
model mpg = foreign;
run;

* Exporting a SAS file as a sas file;
```

```
data lib1.auto1;
set auto1;
run;

* Exporting a SAS file as a csv file;
proc export data= auto1
outfile = "C:\Econometrics\Data\auto1.csv"
dbms=csv replace; putnames=yes;
run;

* Exporting a SAS file as an Excel file;
proc export data=auto1
outfile= "C:\Econometrics\Data\auto1.xls"
dbms=excel replace;
run;
```

Obs	make	price	mpg	repairs	weight	length	foreign
1	AMC	4099	22	3	2930	186	0
2	AMC	4749	17	3	3350	173	0
3	AMC	3799	22	3	2640	168	0
4	Audi	9690	17	5	2830	189	1
5	Audi	6295	23	3	2070	174	1
6	BMW	9735	25	4	2650	177	1
7	Buick	4816	20	3	3250	196	0
8	Buick	7827	15	4	4080	222	0
9	Buick	5788	18	3	3670	218	0
10	Buick	4453	26	3	2230	170	0

The CONTENTS Procedure

Data Set Name	WORK.AUTO	Observations	26
Member Type	DATA	Variables	7
Engine	V9	Indexes	0
Created	Saturday, April 06, 2013 11:32:33 PM	Observation Length	64
Last Modified	Saturday, April 06, 2013 11:32:33 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_32		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information

Data Set Page Size	8192
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	127
Obs in First Data Page	26
Number of Data Set Repairs	0
Filename	C:\...\auto.sas7bdat
Release Created	9.0301M0
Host Created	W32_7PRO

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
7	foreign	Num	8	BEST12.	BEST32.
6	length	Num	8	BEST12.	BEST32.
1	make	Char	11	$11.	$11.
3	mpg	Num	8	BEST12.	BEST32.
2	price	Num	8	BEST12.	BEST32.
4	repairs	Num	8	BEST12.	BEST32.
5	weight	Num	8	BEST12.	BEST32.

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
price	26	6651.73	3371.12	3299.00	15906.00
mpg	26	20.9230769	4.7575042	14.0000000	35.0000000
repairs	26	3.2692308	0.7775702	2.0000000	5.0000000
weight	26	3099.23	695.0794089	2020.00	4330.00
length	26	190.0769231	18.1701361	163.0000000	222.0000000
foreign	26	0.2692308	0.4523443	0	1.0000000

The UNIVARIATE Procedure
Variable: price

Moments

N	26	Sum Weights	26
Mean	6651.73077	Sum Observations	172945
Std Deviation	3371.11981	Variance	11364448.8
Skewness	1.470727	Kurtosis	1.5346717
Uncorrected SS	1434494797	Corrected SS	284111219
Coeff Variation	50.6803406	Std Error Mean	661.130988

Basic Statistical Measures

Location		Variability	
Mean	6651.731	Std Deviation	3371
Median	5146.500	Variance	11364449
Mode	.	Range	12607
		Interquartile Range	3676

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	10.06114	Pr > |t|	<.0001
Sign	M	13	Pr >= |M|	<.0001
Signed Rank	S	175.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	15906.0
99%	15906.0
95%	14500.0
90%	11385.0
75% Q3	8129.0
50% Median	5146.5
25% Q1	4453.0
10%	3799.0
5%	3667.0
1%	3299.0
0% Min	3299.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
3299	17	9735	6
3667	18	10372	13
3799	1	11385	14
3955	19	14500	15
4082	7	15906	16

The UNIVARIATE Procedure
Variable: mpg

Moments			
N	26	Sum Weights	26
Mean	20.9230769	Sum Observations	544
Std Deviation	4.75750419	Variance	22.6338462

Moments			
Skewness	0.93547297	Kurtosis	1.79270004
Uncorrected SS	11948	Corrected SS	565.846154
Coeff Variation	22.7380715	Std Error Mean	0.93302334

Basic Statistical Measures			
Location		Variability	
Mean	20.92308	Std Deviation	4.75750
Median	21.00000	Variance	22.63385
Mode	22.00000	Range	21.00000
		Interquartile Range	6.00000

Tests for Location: Mu0=0				
Test		Statistic		p Value
Student's t	t	22.42503	Pr > |t|	<.0001
Sign	M	13	Pr >= |M|	<.0001
Signed Rank	S	175.5	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	35
99%	35
95%	29
90%	26
75% Q3	23
50% Median	21
25% Q1	17
10%	15
5%	14
1%	14
0% Min	14

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
14	15	24	24
14	14	25	6
15	12	26	8

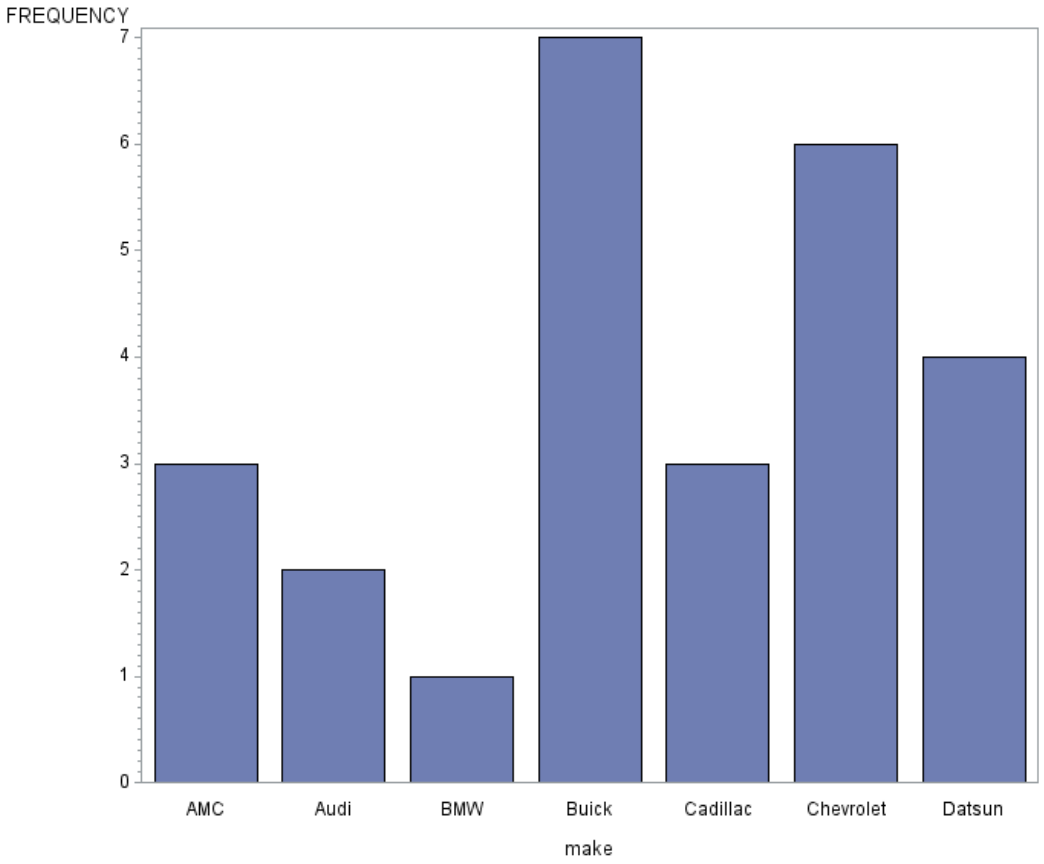
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
16	22	29	17
16	13	35	23

The FREQ Procedure

make	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AMC	3	11.54	3	11.54
Audi	2	7.69	5	19.23
BMW	1	3.85	6	23.08
Buick	7	26.92	13	50.00
Cadillac	3	11.54	16	61.54
Chevrolet	6	23.08	22	84.62
Datsun	4	15.38	26	100.00

foreign	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	19	73.08	19	73.08
1	7	26.92	26	100.00

repairs	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	3	11.54	3	11.54
3	15	57.69	18	69.23
4	6	23.08	24	92.31
5	2	7.69	26	100.00



The SAS System

The CORR Procedure

4 Variables: price mpg weight length

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
price	26	6652	3371	172945	3299	15906
mpg	26	20.92308	4.75750	544.00000	14.00000	35.00000
weight	26	3099	695.07941	80580	2020	4330
length	26	190.07692	18.17014	4942	163.00000	222.00000

Pearson Correlation Coefficients, N = 26
Prob > |r| under H0: Rho=0

	price	mpg	weight	length
price	1.00000	-0.43846	0.55607	0.43604
		0.0251	0.0032	0.0260

Pearson Correlation Coefficients, N = 26
Prob > |r| under H0: Rho=0

	price	mpg	weight	length
mpg	-0.43846	1.00000	-0.80816	-0.76805
	0.0251		<.0001	<.0001
weight	0.55607	-0.80816	1.00000	0.90654
	0.0032	<.0001		<.0001
length	0.43604	-0.76805	0.90654	1.00000
	0.0260	<.0001	<.0001	

The CORR Procedure
foreign=0

4 Variables: price mpg weight length

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
price	19	6484	3768	123199	3299	15906
mpg	19	19.78947	4.03566	376.00000	14.00000	29.00000
weight	19	3348	627.17691	63610	2110	4330
length	19	195.42105	17.96390	3713	163.00000	222.00000

Pearson Correlation Coefficients, N = 19
Prob > |r| under H0: Rho=0

	price	mpg	weight	length
price	1.00000	-0.52852	0.74972	0.52504
		0.0200	0.0002	0.0210
mpg	-0.52852	1.00000	-0.86236	-0.77040
	0.0200		<.0001	0.0001
weight	0.74972	-0.86236	1.00000	0.87771
	0.0002	<.0001		<.0001
length	0.52504	-0.77040	0.87771	1.00000
	0.0210	0.0001	<.0001	

The CORR Procedure
foreign=1

4 Variables: price mpg weight length

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
price	7	7107	2102	49746	4589	9735
mpg	7	24.00000	5.50757	168.00000	17.00000	35.00000
weight	7	2424	325.15930	16970	2020	2830
length	7	175.57143	8.46280	1229	165.00000	189.00000

Pearson Correlation Coefficients, N = 7
Prob > |r| under H0: Rho=0

	price	mpg	weight	length
price	1.00000	-0.64108	0.88279	0.85397
		0.1207	0.0085	0.0144
mpg	-0.64108	1.00000	-0.71010	-0.81171
	0.1207		0.0738	0.0266
weight	0.88279	-0.71010	1.00000	0.87537
	0.0085	0.0738		0.0098
length	0.85397	-0.81171	0.87537	1.00000
	0.0144	0.0266	0.0098	

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: mpg

Number of Observations Read	26
Number of Observations Used	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	378.69701	126.23234	14.84	<.0001
Error	22	187.14915	8.50678		
Corrected Total	25	565.84615			

Root MSE	2.91664	R-Square	0.6693
Dependent Mean	20.92308	Adj R-Sq	0.6242

Coeff Var	13.93982

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	44.96858	9.32268	4.82	<.0001
weight	1	-0.00501	0.00219	-2.29	0.0320
length	1	-0.04306	0.07693	-0.56	0.5813
foreign	1	-1.26921	1.63213	-0.78	0.4451

The GLM Procedure

Class Level Information

Class	Levels	Values
foreign	2	0 1

Number of Observations Read	26
Number of Observations Used	26

The SAS System

The GLM Procedure

Dependent Variable: mpg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	90.6882591	90.6882591	4.58	0.0427
Error	24	475.1578947	19.7982456		
Corrected Total	25	565.8461538			

R-Square	Coeff Var	Root MSE	mpg Mean
0.160270	21.26610	4.449522	20.92308

Source	DF	Type I SS	Mean Square	F Value	Pr > F
foreign	1	90.68825911	90.68825911	4.58	0.0427

Source	DF	Type III SS	Mean Square	F Value	Pr > F
foreign	1	90.68825911	90.68825911	4.58	0.0427