

STAT 415/615

Regression

READING DATA INTO SAS SIMPLE STATISTICS

```
/* ----- DATA statement - importing data into SAS ----- */

/* Just read a SAS data file */

DATA REST; SET "C:\Users\Michael\Teach\520\SAS\rest.sas7bdat"; run;
/* OR: use PROC IMPORT to read a text, Excel, Access, dBase file */

PROC IMPORT OUT=REST
DATAFILE= "C:\Users\Michael\Teach\520\SAS\restaurants.xls"
DBMS=EXCEL2000 REPLACE;
GETNAMES=YES;
RUN;

/* OR: include a *.dat file - but it our file consists of two pieces */

DATA REST1;
infile "C:\Users\Michael\Teach\520\SAS\rest.dat" firstobs=3 obs=281;
input ROW ID LOOK SALES NEWCAP VALUE COSTGOOD WAGES ADS TYPE;

DATA REST2;
infile "C:\Users\Michael\Teach\520\SAS\rest.dat" firstobs=286 obs=564;
input ROW SEATS OWNER FT PT SIZE;

DATA REST; MERGE REST1 REST2; run;

/* ----- What's in the file? ----- */
PROC CONTENTS DATA=REST; RUN;

PROC MEANS DATA=REST; RUN;

/* ----- Descriptive Statistics ----- */
PROC UNIVARIATE DATA=REST NORMAL PLOT;
VAR SALES; RUN;

/* ----- Scatterplots ----- */
PROC PLOT DATA=REST;
PLOT SALES * WAGES; RUN;
```

The SAS System

14:58 Friday, August 27, 2014 6

The CONTENTS Procedure

Data Set Name: WORK.REST

Member Type: DATA

Engine: V8

Created: 15:31 Friday, August 27, 1993

Observations: 279

Variables: 15

Indexes: 0

Observation Length: 120

Last Modified: 15:31 Friday, August 27, 1993
 Protection:
 Data Set Type:
 Label:

Deleted Observations: 0
 Compressed: NO
 Sorted: NO

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos
9	ADS	Num	8	64
7	COSTGOOD	Num	8	48
13	FT	Num	8	96
2	ID	Num	8	8
3	LOOK	Num	8	16
5	NEWCAP	Num	8	32
12	OWNER	Num	8	88
14	PT	Num	8	104
1	ROW	Num	8	0
4	SALES	Num	8	24
11	SEATS	Num	8	80
15	SIZE	Num	8	112
10	TYPE	Num	8	72
6	VALUE	Num	8	40
8	WAGES	Num	8	56

The SAS System 14:58 Friday, August 27, 2014 7

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
ROW	279	140.0000000	80.6845710	1.0000000	279.0000000
ID	279	140.0000000	80.6845710	1.0000000	279.0000000
LOOK	278	3.3597122	1.4294842	1.0000000	7.0000000
SALES	254	332.5629921	650.5393156	0	8064.00
NEWCAP	224	12.9196429	28.1304260	0	300.0000000
VALUE	240	349.1166667	884.9163654	0	12000.00
COSTGOOD	237	45.1308017	13.3427669	0	99.0000000
WAGES	235	25.2510638	10.8860036	0	85.0000000
ADS	235	3.9106383	3.9196652	0	25.0000000
TYPE	267	1.8651685	0.8342888	1.0000000	3.0000000
SEATS	268	71.5410448	62.1460328	0	550.0000000
OWNER	269	2.1115242	0.9436636	1.0000000	3.0000000
FT	265	8.1018868	18.1504914	0	250.0000000
PT	266	12.7293233	14.0761422	0	100.0000000
SIZE	263	1.6844106	0.8074228	1.0000000	3.0000000

The SAS System 14:58 Friday, August 27, 2014 8

The UNIVARIATE Procedure
 Variable: SALES

Moments

N	254	Sum Weights	254
Mean	332.562992	Sum Observations	84471
Std Deviation	650.539316	Variance	423201.401
Skewness	8.41196889	Kurtosis	88.2236891
Uncorrected SS	135161883	Corrected SS	107069954
Coeff Variation	195.613863	Std Error Mean	40.818467

Basic Statistical Measures

Location		Variability	
Mean	332.5630	Std Deviation	650.53932

Median	200.0000	Variance	423201
Mode	250.0000	Range	8064
		Interquartile Range	297.00000

Tests for Location: $\mu_0=0$

Test	-Statistic-	-----p Value-----
Student's t	t 8.147366	Pr > t <.0001
Sign	M 126.5	Pr >= M <.0001
Signed Rank	S 16065.5	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	8064
99%	3450
95%	1000
90%	647
75% Q3	382
50% Median	200
25% Q1	85
10%	50
5%	30
1%	2
0% Min	0

The SAS System 14:58 Friday, August 27, 2014 9

The UNIVARIATE Procedure
Variable: SALES

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
0	203	1200	55
2	182	1600	191
2	178	3450	111
8	255	4700	57
10	237	8064	219

Missing Values

Missing Value	Count	-----Percent Of-----	
		All Obs	Missing Obs
.	25	8.96	100.00

Histogram

8250+*

4250+

250+*

Figure 1 is a scatter plot showing the distribution of normalized residuals for the 1995-1996 season. The x-axis is labeled "Normalized Residuals" and ranges from -2 to +2. The y-axis is labeled "Frequency" and ranges from 250 to 8250. The plot shows a large number of residuals clustered near zero, with a few outliers at higher frequencies.

Plot of SALES*WAGES. Legend: A = 1 obs, B = 2 obs, etc.

```
SALES ,
8000 ^
,
,
< cropped >
,
5000 ^
,
,
A
,
,
< cropped >
,
,
,
,
,
,
A
,
,
A A
```



NOTE: 56 obs had missing values.

Description of Restaurants data

This survey was conducted in 1980 by the University of Wisconsin Small Business Development Center. The purpose was "to develop data that would be useful to small business counselors in advising managers as to how to effectively plan and operate their small restaurants".

Nineteen of Wisconsin's counties were selected for study. Lists of restaurants were drawn up from telephone directories and these were sampled in proportion to the population of the county. A sample of 1000 restaurants yielded 279 usable responses. The data base thus consists of 279 summaries of the returned questionnaires. Some respondents failed to present complete answers. Thus, there are a number of missing values, which are coded by "."

List of variables

No	Variable	Description
1	ID	Identification number (1-279)
2	OUTLOOK	Values 1,2,3,4,5,6 denoting from very unfavorable to very favorable
3	SALES	Gross 1979 sales in \$1000s
4	NEWCAP	New capital invested in 1979, in \$1000s
5	VALUE	Estimated market value of the business, in \$1000s
6	COSTGOOD	Cost of goods sold as a percentage of sales
7	WAGES	Wages as a percentage of sales
8	ADS	Advertising as a percentage of sales
9	TYPEFOOD	1 = fast food, 2 = supper club, 3 = other
10	SEATS	Number of seats in dining area
11	OWNER	1 = sole proprietorship, 2 = partnership, 3 = corporation
12	FT.EMPL	Number of full-time employees
13	PT.EMPL	Number of part-time employees
14	SIZE	Size of restaurant: 1 = 1 to 9.5 full-time equivalent employees (f.t.e.e.) 2 = 10 to 20 f.t.e.e., 3 = over 20 f.t.e.e., where f.t.e.e. equals (number of full time) + (1/2)(number of part time)