

STA442s04 Overheads: Set One

```
/res/jbrunner/442s04 > ls
furnace.raw      heat1.sas
/res/jbrunner/442s04 > sas heat1
/res/jbrunner/442s04 > ls
furnace.raw      heat1.log      heat1.lst      heat1.sas
/res/jbrunner/442s04 >

/res/jbrunner/442s04 > cat heat1.sas

/***** heat1.sas *****/

options linesize=79 pagesize=35 noovp formdlim='_';
title 'Read and Describe Furnace Data';

proc format;
    value frnfmt 1= 'Forced air' 2= 'Gravity' 3= 'Forced water';
    value shfmt 1= 'Round' 2= 'Square' 3= 'Rectangular';
    value lnfmt 0= 'Unlined' 1= 'Tile' 2= 'Metal';
    value hfmt 1= 'Ranch' 2= 'Two-story' 3= 'tri-level'
              4= 'Bi-level' 5= '1.5 stories';
    value dfmt 2= 'EVD' 1= 'TVD'; /* direction reversed */
    value catfmt 1= 'Ranch' 2= 'Two Story' 3 = 'Other';

data warm;
    infile 'furnace.raw';
    input id typfurn area shape height liner house age
           dampin dampout damper;

    label id = 'House Identification Number';
    label typfurn = 'Type of furnace';
    label area = 'Chimney area';
    label shape = 'Chimney shape';
    label height = 'Chimney height in feet';
    label liner = 'Type of Chimney liner';
    label house = 'Type of house';
    label age = 'House age in yrs (99=99+)';
    label damper = 'Type of damper';
    label dampin = 'Energy consumpt with damper in';
    label dampout = 'Energy consumpt with damper out';

    format typfurn frnfmt.;
    format shape shfmt.;
    format liner lnfmt.;
    format house hfmt.;
    format damper dfmt.;
```

```

/** SAS doesn't like numeric missing value codes. A period . Is
    best for missing. However, suppose 999=missing .... */

    if typfurn eq 999 then typfurn = . ;
    if shape eq 999 then shape = . ;

/***** Creating New Variables *****/

    consume = (dampin+dampout)/2;
    label consume = 'Aver Energy Consumpt';

    diff=dampout-dampin;
    label diff = 'consumpt w/ damper out minus in';

    if house=. then housecat=.;
    else if house=1 then housecat=1;
    else if house=2 then housecat=2;
    else housecat=3;
    label housecat = 'Recoded House Type';
    format housecat catfmt.;

proc freq;
    title2 'Frequency Distributions of everything';
    tables _all_; /* Even id and continuous variables -- Don't print this!*/

proc freq;
    title2 'Frequency Distributions of categorical variables';
    tables typfurn shape liner house housecat damper;

proc freq;
    title2 'Always test new categorical variables';
    tables house*housecat / norow nocol nopercnt missing;

proc means;
    title2 'Quantitative vars';
    var area height age -- dampout consume diff;
    /* Equivalent to area height age dampin dampout consume diff. Single
       dash only works with numbered lists, like item1-item50 */

/* More detailed look at the quantitative variables, with proc univariate */

proc univariate plot;
    title2 'PROC UNIVARIATE on Quantitative vars';
    var area height age -- dampout consume diff;

/* You can get crude statistical graphics -- maybe okay for a first look.
   There is a good variety of chart types. */
proc chart ;
    title2 'Vertical Bar Chart of Chimney Shape';
    vbar shape/discrete type=percent;

```

/res/jbrunner/442s04 > cat heat1.lst

Cutting out the first 26 pages ...

Read and Describe Furnace Data 27
Frequency Distributions of categorical variables
13:49 Saturday, December 27, 2003

Type of furnace

TYPFURN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Forced air	76	84.4	76	84.4
Gravity	7	7.8	83	92.2
Forced water	7	7.8	90	100.0

Chimney shape

SHAPE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Round	39	43.8	39	43.8
Square	32	36.0	71	79.8
Rectangular	18	20.2	89	100.0

Frequency Missing = 1

Read and Describe Furnace Data 28
Frequency Distributions of categorical variables
13:49 Saturday, December 27, 2003

Type of Chimney liner

LINER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Unlined	24	27.0	24	27.0
Tile	40	44.9	64	71.9
Metal	25	28.1	89	100.0

Frequency Missing = 1

Type of house

HOUSE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Ranch	38	42.2	38	42.2
Two-story	40	44.4	78	86.7
tri-level	3	3.3	81	90.0
Bi-level	6	6.7	87	96.7
1.5 stories	3	3.3	90	100.0

Recoded House Type

HOUSECAT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Ranch	38	42.2	38	42.2
Two Story	40	44.4	78	86.7
Other	12	13.3	90	100.0

Type of damper

DAMPER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
TVD	40	44.4	40	44.4
EVD	50	55.6	90	100.0

TABLE OF HOUSE BY HOUSECAT

HOUSE (Type of house)	HOUSECAT (Recoded House Type)			Total
Frequency	Ranch	Two Stor y	Other	
Ranch	38	0	0	38
Two-story	0	40	0	40
tri-level	0	0	3	3
Bi-level	0	0	6	6
1.5 stories	0	0	3	3
Total	38	40	12	90

Read and Describe Furnace Data
Quantitative vars

31

13:49 Saturday, December 27, 2003

Variable	Label	N	Mean	Std Dev
AREA	Chimney area	89	62.5617978	32.5307390
HEIGHT	Chimney height in feet	90	21.9666667	5.9254735
AGE	House age in yrs (99=99+)	90	38.5666667	31.0932089
DAMPIN	Energy consumpt with damper in	90	10.0384444	2.8679903
DAMPOUT	Energy consumpt with damper out	90	10.8131111	3.0884073
CONSUME	Aver Energy Consumpt	90	10.4257778	2.9641170
DIFF	consumpt w/ damper out minus in	90	0.7746667	0.6191099

Variable	Label	Minimum	Maximum
AREA	Chimney area	28.0000000	168.0000000
HEIGHT	Chimney height in feet	14.0000000	39.0000000
AGE	House age in yrs (99=99+)	1.0000000	99.0000000
DAMPIN	Energy consumpt with damper in	2.9700000	18.2600000
DAMPOUT	Energy consumpt with damper out	3.2000000	20.5500000
CONSUME	Aver Energy Consumpt	3.0850000	19.4050000
DIFF	consumpt w/ damper out minus in	-0.8700000	3.9800000

PROC UNIVARIATE on Quantitative vars

13:49 Saturday, December 27, 2003

Univariate Procedure

Variable=AREA Chimney area

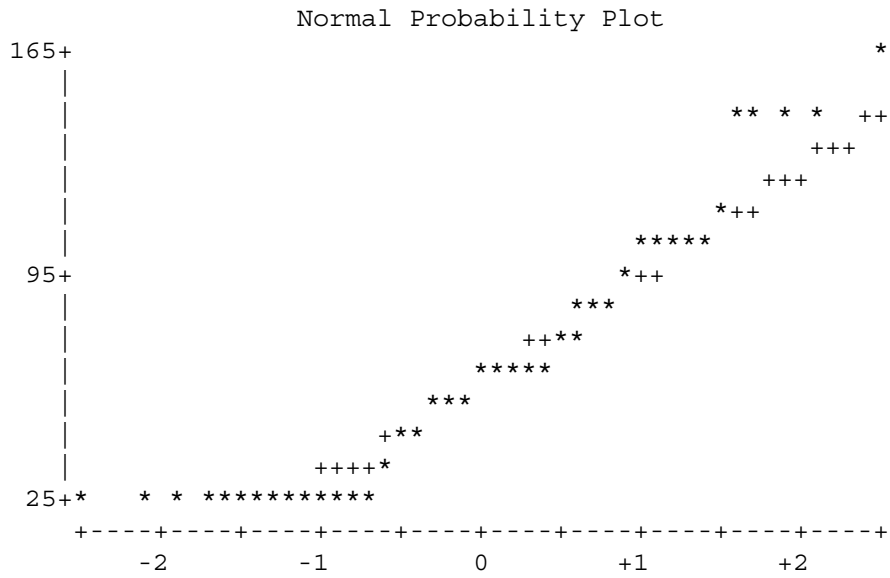
Moments

N	89	Sum Wgts	89
Mean	62.5618	Sum	5568
Std Dev	32.53074	Variance	1058.249
Skewness	1.06275	Kurtosis	0.914776
USS	441470	CSS	93125.91
CV	51.99777	Std Mean	3.448251
T:Mean=0	18.14305	Pr> T	0.0001
Num ^= 0	89	Num > 0	89
M(Sign)	44.5	Pr>= M	0.0001
Sgn Rank	2002.5	Pr>= S	0.0001

Univariate Procedure

Variable=AREA

Chimney area



Same thing for the other quantitative variables. Skip to diff.

Univariate Procedure

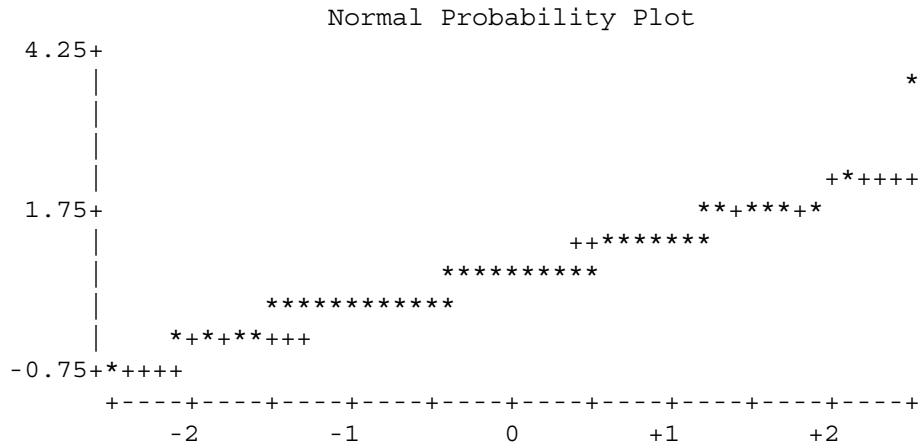
Variable=DIFF

consumpt w/ damper out minus in

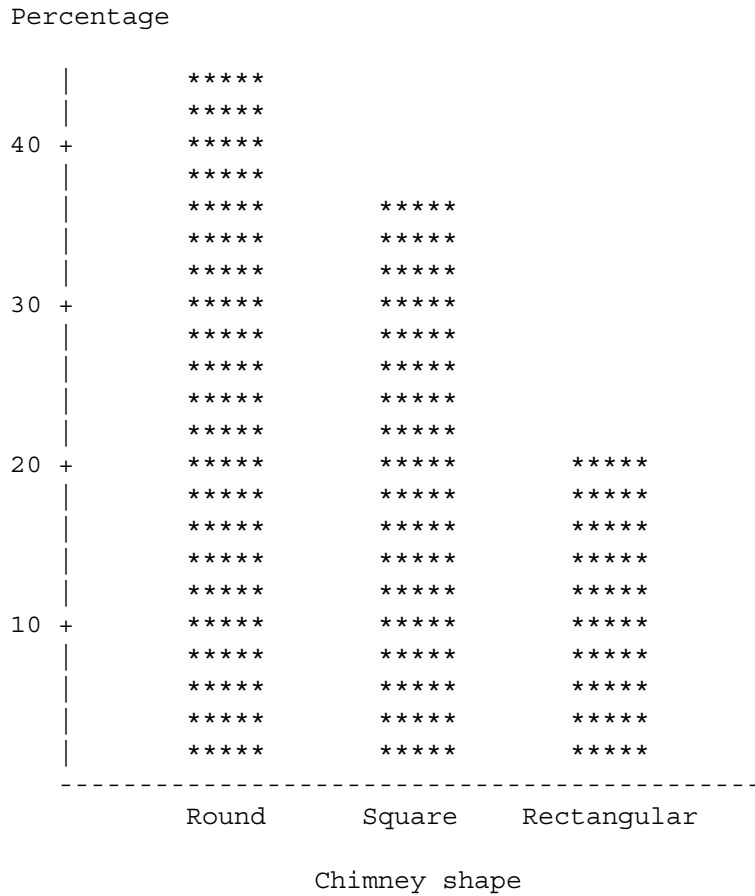
Moments

N	90	Sum Wgts	90
Mean	0.774667	Sum	69.72
Std Dev	0.61911	Variance	0.383297
Skewness	1.542418	Kurtosis	7.600262
USS	88.1232	CSS	34.11344
CV	79.91952	Std Mean	0.06526
T:Mean=0	11.87048	Pr> T	0.0001
Num ^= 0	90	Num > 0	85
M(Sign)	40	Pr>= M	0.0001
Sgn Rank	1948.5	Pr>= S	0.0001

Quantiles (Def=5)



Read and Describe Furnace Data 58
 Vertical Bar Chart of Chimney Shape
 13:49 Saturday, December 27, 2003



/res/jbrunner/442s04 >