An Old Assignment on Elementary tests

Using the FURNACE data, carry out elementary significance tests to answer these questions. There is not necessarily a separate SAS procedure for each question. In each of your analyses, there is one independent variable and one dependent variable.

1. If you observe the shape of a house's chimney, does that improve your ability to predict what type of furnace the house has?

2. Is there more average energy consumption with the vent damper active, or inactive?

3. Is there a tendency for houses that consume lots of energy with the vent damper inactive to also consume a lot of energy with the vent damper active?

4. Does energy consumption depend on type of vent damper? Your dependent variable should be the *difference* of two variables in the raw data file.

5. Does average amount of energy consumption with vent damper inactive depend on type of chimney liner? If the answer is yes, which means are different from each other?

6. Do different kinds of house tend to have different types of chimney?

7. Make a new variable that is house age at or below the median versus house age above the median (refer to earlier printouts). Using this variable, do older houses tend to use more energy with the vent damper inactive? Why did I ask for this instead of a simple regression?

8. What proportion of the variation in chimney height is explained by chimney type?

9. What proportion of the variation in energy consumption with vent damper active is explained by type of damper? Is it significant? What do you conclude? Did you do follow-up tests? Why or why not?

10. What proportion of the variation in energy consumption with vent damper active is explained by chimney height? Is it significant? What do you conclude?

11. Consider the *difference* in energy consumption between damper active and damper inactive. What proportion of the variation in this variable is explained by furnace type? Is it significant? What do you conclude?

```
/* 1.  If you observe the shape of a house's chimney, does that improve your
ability to predict what type of furnace the house has? */

proc freq;
     tables shape*typfurn / nopercent nocol chisq expected;
```

TABLE OF SHAPE BY TYPFURN

SHAPE(Chimney shape) TYPFURN(Type of furnace)

Frequency / Expected / Row Pct	Forced air	Gravity	Forced water	Total
Round	37 33.303 94.87	1 3.0674 2.56	1 2.6292 2.56	39
Square	25 27.326 78.13	5 2.5169 15.63	2 2.1573 6.25	32
Rectangular	14 15.371 77.78	1 1.4157 5.56	3 1.2135 16.67	18
Total	76	7	6	89

Frequency Missing = 1

STATISTICS FOR TABLE OF SHAPE BY TYPFURN

Statistic	DF	Value	Prob
Chi-Square	4	8.347	0.080
Likelihood Ratio Chi-Square	4	7.848	0.097
Mantel-Haenszel Chi-Square	1	4.516	0.034
Phi Coefficient		0.306	
Contingency Coefficient		0.293	
Cramer's V		0.217	

Effective Sample Size = 89
Frequency Missing = 1
WARNING: 67% of the cells have expected counts less
         than 5. Chi-Square may not be a valid test.

```
/* 2.  Is there more average energy consumption with the vent damper active,
or inactive?  */

proc means   mean std n t prt; /* Did this in heat2.sas */
     var diff;
```

```
                                Furnace Data                                3
                                                13:46 Sunday, January 4, 2004

          Analysis Variable : DIFF consumpt w/ damper out minus in

            Mean       Std Dev   N             T  Prob>|T|
          ----------------------------------------------------
          0.7746667     0.6191099  90    11.8704824    0.0001
          ----------------------------------------------------
```

```
/* 3.  Is there a tendency for houses that consume lots of energy with the
vent damper inactive to also consume a lot of energy with the vent damper
active?  */

proc corr nosimple;      /* Did this already */
     var dampin dampout;
```

```
                                Furnace Data                                4
                                                13:46 Sunday, January 4, 2004

                             Correlation Analysis

                       2 'VAR' Variables:  DAMPIN   DAMPOUT


   Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 90

                                              DAMPIN           DAMPOUT

     DAMPIN                                  1.00000           0.98111
     Energy consumpt with damper in           0.0               0.0001

     DAMPOUT                                 0.98111           1.00000
     Energy consumpt with damper out          0.0001            0.0
```

```
/* 4.  Does energy consumption depend on type of vent damper? Your dependent
variable should be the difference of two variables in the raw data file.  */

proc ttest;            /* Did this already */
     class damper;
     var diff ;
```

```
                                Furnace Data                                5
                                                13:46 Sunday, January 4, 2004

                              TTEST PROCEDURE

Variable: DIFF         consumpt w/ damper out minus in

DAMPER       N         Mean      Std Dev    Std Error      Minimum      Maximum
-------------------------------------------------------------------------------
   TVD      40   0.66150000   0.51063334   0.08073822  -0.38000000   2.29000000
   EVD      50   0.86520000   0.68545007   0.09693728  -0.87000000   3.98000000

Variances        T       DF    Prob>|T|
---------------------------------------
Unequal    -1.6147     87.6      0.1100
Equal      -1.5636     88.0      0.1215

For H0: Variances are equal, F' = 1.80    DF = (49,39)    Prob>F' = 0.0596
```

```
/* 5.  Does average amount of energy consumption with vent damper inactive
depend on type of chimney liner? If the answer is yes, which means are
different from each other?  */

proc glm;
     class liner;
     model dampout=liner;
     means liner;
     means liner / bon scheffe tukey;
```

```
                                Furnace Data                                6

                        General Linear Models Procedure
                            Class Level Information

                     Class    Levels    Values

                     LINER         3    Metal Tile Unlined


                    Number of observations in data set = 90
```

NOTE: Due to missing values, only 89 observations can be used in this analysis.

General Linear Models Procedure

Dependent Variable: DAMPOUT Energy consumpt with damper out

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	77.36449661	38.68224831	4.49	0.0140
Error	86	741.09641350	8.61740016		
Corrected Total	88	818.46091011			

R-Square	C.V.	Root MSE	DAMPOUT Mean
0.094524	27.30365	2.9355409	10.751461

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LINER	2	77.36449661	38.68224831	4.49	0.0140

Source	DF	Type III SS	Mean Square	F Value	Pr > F
LINER	2	77.36449661	38.68224831	4.49	0.0140

General Linear Models Procedure

Level of LINER	N	DAMPOUT Mean	DAMPOUT SD
Metal	25	9.2884000	2.51587241
Tile	40	11.1567500	2.71503637
Unlined	24	11.6000000	3.62179514

General Linear Models Procedure

Tukey's Studentized Range (HSD) Test for variable: DAMPOUT

NOTE: This test controls the type I experimentwise error rate.

Alpha= 0.05 Confidence= 0.95 df= 86 MSE= 8.6174
Critical Value of Studentized Range= 3.373

Comparisons significant at the 0.05 level are indicated by '***'.

LINER Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
Unlined - Tile	-1.3645	0.4433	2.2510	
Unlined - Metal	0.3108	2.3116	4.3124	***
Tile - Unlined	-2.2510	-0.4433	1.3645	
Tile - Metal	0.0834	1.8684	3.6533	***
Metal - Unlined	-4.3124	-2.3116	-0.3108	***
Metal - Tile	-3.6533	-1.8684	-0.0834	***

```
/* 6.  Do different kinds of house tend to have different types of
chimney? */

proc freq;
     tables (house housecat)*shape / nopercent nocol chisq expected;
```

TABLE OF HOUSE BY SHAPE

HOUSE(Type of house) SHAPE(Chimney shape)

Frequency / Expected / Row Pct	Round	Square	Rectangular	Total
Ranch	22	13	3	38
	16.652	13.663	7.6854	
	57.89	34.21	7.89	
Two-story	8	18	13	39
	17.09	14.022	7.8876	
	20.51	46.15	33.33	
tri-level	3	0	0	3
	1.3146	1.0787	0.6067	
	100.00	0.00	0.00	
Bi-level	6	0	0	6
	2.6292	2.1573	1.2135	
	100.00	0.00	0.00	
1.5 stories	0	1	2	3
	1.3146	1.0787	0.6067	
	0.00	33.33	66.67	
Total	39	32	18	89

Frequency Missing = 1

STATISTICS FOR TABLE OF HOUSE BY SHAPE

Statistic	DF	Value	Prob
Chi-Square	8	29.941	0.001
Likelihood Ratio Chi-Square	8	34.625	0.001
Mantel-Haenszel Chi-Square	1	0.844	0.358
Phi Coefficient		0.580	
Contingency Coefficient		0.502	
Cramer's V		0.410	

```
Effective Sample Size = 89
Frequency Missing = 1
WARNING: 60% of the cells have expected counts less
         than 5. Chi-Square may not be a valid test.
```

TABLE OF HOUSECAT BY SHAPE

HOUSECAT(Recoded House Type) SHAPE(Chimney shape)

Frequency / Expected / Row Pct	Round	Square	Rectangular	Total
Ranch	22	13	3	38
	16.652	13.663	7.6854	
	57.89	34.21	7.89	
Two Story	8	18	13	39
	17.09	14.022	7.8876	
	20.51	46.15	33.33	
Other	9	1	2	12
	5.2584	4.3146	2.427	
	75.00	8.33	16.67	
Total	39	32	18	89

Frequency Missing = 1

STATISTICS FOR TABLE OF HOUSECAT BY SHAPE

Statistic	DF	Value	Prob
Chi-Square	4	19.167	0.001
Likelihood Ratio Chi-Square	4	21.129	0.001
Mantel-Haenszel Chi-Square	1	1.372	0.242
Phi Coefficient		0.464	
Contingency Coefficient		0.421	
Cramer's V		0.328	

```
Effective Sample Size = 89
Frequency Missing = 1
WARNING: 22% of the cells have expected counts less
         than 5. Chi-Square may not be a valid test.
```

```
/* 7.  Make a new variable that is house age at or below the median versus
house age above the median (refer to earlier printouts). Using this variable,
do older houses tend to use more energy with the vent damper inactive? Why did
I ask for this instead of a simple regression?  */

proc ttest;
     class agecat;
     var dampout ;
```

This was at the beginning of heat3.sas

```
/* Still part of the data step -- needed for Question 7  */
     if age = . then agecat = . ;
        else if age <= 30 then agecat = 0;
        else agecat = 1;
     label agecat = 'House Age Above Median?';
proc format;
           value fmt 0 = 'At or Below Med'  1 = 'Above Median';
proc plot; /* Just checking */
     plot age * agecat;
```

```
                                  Furnace Data                                13
                                                 13:46 Sunday, January 4, 2004

                                TTEST PROCEDURE

Variable: DAMPOUT      Energy consumpt with damper out

AGECAT       N         Mean      Std Dev    Std Error      Minimum      Maximum
-------------------------------------------------------------------------------
     0      46  10.14130435   2.42441710   0.35746080   3.20000000  14.77000000
     1      44  11.51545455   3.55010216   0.53519804   4.29000000  20.55000000

Variances        T       DF    Prob>|T|
---------------------------------------
Unequal    -2.1351     75.6      0.0360
Equal      -2.1527     88.0      0.0341

For H0: Variances are equal, F' = 2.14    DF = (43,45)    Prob>F' = 0.0126
```

```
/* 8.  What proportion of the variation in chimney height is explained by
chimney type?  */

proc glm;
     class shape;
     model height=shape;
     means shape;
     means shape / bon scheffe tukey;
```

General Linear Models Procedure
Class Level Information

Class	Levels	Values
SHAPE	3	Rectangular Round Square

Number of observations in data set = 90

NOTE: Due to missing values, only 89 observations can be used in this analysis.

General Linear Models Procedure

Dependent Variable: HEIGHT Chimney height in feet

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1140.2953640	570.1476820	25.55	0.0001
Error	86	1919.3450855	22.3179661		
Corrected Total	88	3059.6404494			

R-Square	C.V.	Root MSE	HEIGHT Mean
0.372689	21.59491	4.7241895	21.876404

```
/* 9.  What proportion of the variation in energy consumption with vent damper
active is explained by type of damper? Is it significant? What do you
conclude? Did you do follow-up tests? Why or why not?  */

proc glm;
     class damper;
     model dampin=damper;
     means damper;
```

General Linear Models Procedure

Dependent Variable: DAMPIN Energy consumpt with damper in

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.22983472	1.22983472	0.15	0.7013
Error	88	730.82794750	8.30486304		
Corrected Total	89	732.05778222			

R-Square	C.V.	Root MSE	DAMPIN Mean
0.001680	28.70779	2.8818159	10.038444

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DAMPER	1	1.22983472	1.22983472	0.15	0.7013

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DAMPER	1	1.22983472	1.22983472	0.15	0.7013

General Linear Models Procedure

Level of DAMPER	N	DAMPIN Mean	DAMPIN SD
EVD	50	10.1430000	2.76701950
TVD	40	9.9077500	3.01986796

```
 /* 10.  What proportion of the variation in energy consumption with vent
damper active is explained by chimney height? Is it significant? What do you
conclude?  */

proc reg;
     model dampin = height;
```

```
                                Furnace Data                                23
                                                  13:46 Sunday, January 4, 2004

Model: MODEL1
Dependent Variable: DAMPIN     Energy consumpt with damper in

                           Analysis of Variance

                              Sum of          Mean
       Source          DF     Squares        Square      F Value       Prob>F

       Model            1    18.08789      18.08789        2.229       0.1390
       Error           88   713.96989       8.11329
       C Total         89   732.05778

           Root MSE       2.84838     R-square       0.0247
           Dep Mean      10.03844     Adj R-sq       0.0136
           C.V.          28.37476

                        Parameter Estimates

                    Parameter      Standard    T for H0:
     Variable  DF      Estimate         Error   Parameter=0    Prob > |T|

     INTERCEP   1      8.367200    1.15886577         7.220        0.0001
     HEIGHT     1      0.076081    0.05095427         1.493        0.1390
```

```
/* 11.  Consider the difference in energy consumption between damper active
and damper inactive. What proportion of the variation in this variable is
explained by furnace type? Is it significant? What do you conclude?  */

proc glm;
     class typfurn;
     model diff=typfurn;
     means typfurn;
     means typfurn / bon scheffe tukey;
```

```
                                Furnace Data                                25
                                              13:46 Sunday, January 4, 2004

                        General Linear Models Procedure

Dependent Variable: DIFF   consumpt w/ damper out minus in
                                   Sum of            Mean
Source                     DF          Squares          Square  F Value    Pr > F

Model                       2       0.69834432      0.34917216     0.91    0.4067

Error                      87      33.41509568      0.38408156

Corrected Total            89      34.11344000

                  R-Square             C.V.        Root MSE          DIFF Mean

                  0.020471         80.00127       0.6197431          0.7746667
```

```
                                Furnace Data                                26
                                              13:46 Sunday, January 4, 2004

                        General Linear Models Procedure

              Level of              -------------DIFF------------
              TYPFURN          N       Mean              SD

              Forced  air     76     0.79434211       0.63196695
              Forced  water    7     0.47714286       0.57398357
              Gravity          7     0.85857143       0.49744107
```

```
/********************** heat3.sas ****************************/
title2 'Old assignment on choice of elementary tests';

%include 'heatread.sas'; /* Basically the data step from heat1.sas */

/* Still part of the data step -- needed for Question 7  */
     if age = . then agecat = . ;
        else if age <= 30 then agecat = 0;
        else agecat = 1;
     label agecat = 'House Age Above Median?';
proc format;
           value fmt 0 = 'At or Below Med'  1 = 'Above Median';
proc plot; /* Just checking */
     plot age * agecat;


/* 1.  If you observe the shape of a house's chimney, does that improve your
ability to predict what type of furnace the house has? */

proc freq;
     tables shape*typfurn / nopercent nocol chisq expected;

/* 2.  Is there more average energy consumption with the vent damper active,
or inactive?  */

proc means   mean std n t prt; /* Did this in heat2.sas */
     var diff;

/* 3.  Is there a tendency for houses that consume lots of energy with the
vent damper inactive to also consume a lot of energy with the vent damper
active?  */

proc corr nosimple;      /* Did this already */
     var dampin dampout;

/* 4.  Does energy consumption depend on type of vent damper? Your dependent
variable should be the difference of two variables in the raw data file.  */

proc ttest;            /* Did this already */
     class damper;
     var diff ;

/* 5.  Does average amount of energy consumption with vent damper inactive
depend on type of chimney liner? If the answer is yes, which means are
different from each other?  */

proc glm;
     class liner;
     model dampout=liner;
     means liner;
     means liner / bon scheffe tukey;
```

```
/* 6.  Do different kinds of house tend to have different types of
chimney? */

proc freq;
     tables (house housecat)*shape / nopercent nocol chisq expected;

/* 7.  Make a new variable that is house age at or below the median versus
house age above the median (refer to earlier printouts). Using this variable,
do older houses tend to use more energy with the vent damper inactive? Why did
I ask for this instead of a simple regression?  */

proc ttest;
     class agecat;
     var dampout ;

/* 8.  What proportion of the variation in chimney height is explained by
chimney type?  */

proc glm;
     class shape;
     model height=shape;
     means shape;
     means shape / bon scheffe tukey;

/* 9.  What proportion of the variation in energy consumption with vent damper
active is explained by type of damper? Is it significant? What do you
conclude? Did you do follow-up tests? Why or why not?  */

proc glm;
     class damper;
     model dampin=damper;
     means damper;


/* 10.  What proportion of the variation in energy consumption with vent
damper active is explained by chimney height? Is it significant? What do you
conclude?  */

proc reg;
     model dampin = height;


/* 11.  Consider the difference in energy consumption between damper active
and damper inactive. What proportion of the variation in this variable is
explained by furnace type? Is it significant? What do you conclude?  */

proc glm;
     class typfurn;
     model diff=typfurn;
     means typfurn;
     means typfurn / bon scheffe tukey;
```